

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Minnafield Examiner #: \_\_\_\_\_ Date: 3/6/02  
 Art Unit: 1645 Phone Number 30 53394 Serial Number: 91004395  
 Mail Box and Bldg/Room Location: CM1-8E12 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Recombinant P37/FlaA as a Diagnostic Reagent  
 Inventors (please provide full names): Robert D. Gilmore Jr; Barbara J.B. Johnson

Earliest Priority Filing Date: Jan 8, 1998

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

SEQ SEARCH → SEQ 1 (1-1665)  
 SEQ 2 (-341)  
 SEQ 3 (1-21)

Point of Contact  
 Mona Smith  
 Technical Information Specialist  
 CM1 6A01  
 Tel: 308-3278

RECEIVED  
 MAR-6-2002  
 (STIC)

Please search attached claims

Reviewed  
 3/26/02

Lyme Disease + Borrelia Burgdorferi

Diagnostic

Serodiagnostic antigen

outer sheath protein of periplasmic flagella

Thanks  
 Minnafield

## STAFF USE ONLY

Searcher: M. Smith  
 Searcher Phone #: \_\_\_\_\_  
 Searcher Location: \_\_\_\_\_  
 Date Searcher Picked Up: \_\_\_\_\_  
 Date Completed: 3/25/02  
 Searcher Prep & Review Time: 60  
 Clerical Prep Time: \_\_\_\_\_  
 Online Time: 60

## Type of Search

NA Sequence (#) 2  
 AA Sequence (#) 1  
 Structure (#) \_\_\_\_\_  
 Bibliographic X  
 Litigation \_\_\_\_\_  
 Fulltext \_\_\_\_\_  
 Patent Family \_\_\_\_\_  
 Other \_\_\_\_\_

## Vendors and cost where applicable

STN \_\_\_\_\_  
 Dialog \_\_\_\_\_  
 Questel/Orbit \_\_\_\_\_  
 Dr.Link \_\_\_\_\_  
 Lexis/Nexis \_\_\_\_\_  
 Sequence Systems \_\_\_\_\_  
 WWW/Internet \_\_\_\_\_  
 Other (specify) \_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 23:36:44 ; Search time 2486.21 Seconds

(Without alignments)  
11034.792 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663

Sequence: 1 atgataatcttttttcaaaa.....atatccttgaagaataatt 1663

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_scs:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_in:\*  
32: em\_htgo\_rod:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1637	98.4	1655	1	BBU62900	U62900 Borrelia bu
2	1210.6	72.8	13271	1	AE001168	AE001168 Borrelia
3	435.8	26.2	10016	1	AE001167	AE001167 Borrelia
4	314	18.9	332	1	BBCHEAGEN	X91907 B. burgdorferi
5	164.4	9.9	2700	1	BBU28962	BBU28962 Borrelia bu
6	108.6	6.5	1140	6	A24079	A24079 T. hyodysent
7	108.6	6.5	1260	1	THFLAAL	X63006 T. hyodysent
8	93.2	5.6	67970	3	PFMALIP3	AL031746 Plasmodiu
9	89.4	5.4	173957	3	AC006986	AC006986 Homo sapi
10	86.8	5.2	110000	2	PFMALIP1_0	AL034557 Plasmodiu
11	84.8	5.1	1128	3	MPU15677	U15677 Myrmecia pl
12	81.2	4.9	242513	2	AC079314	AC079314 Homo sapi
13	81	4.9	175229	2	CNSO1RQ0	AL159141 Homo sapi
14	81	4.9	209157	9	CNSO1DM4	AL031746 Human chr
15	79.6	4.8	179537	2	AC013513	AC013513 Homo sapi
16	79	4.8	2192	3	AF298207	AF298207 Dictyoste
17	78.8	4.7	253305	3	PFMALIP37	AL034559 Plasmodiu
18	78.4	4.7	60604	2	AC023456	AC023456 Homo sapi
19	78	4.7	19124	6	PFAYARIA	L40608 Plasmodiu
20	78	4.7	19124	6	AR089157	AR089157 Sequence
21	77.2	4.6	2718	1	AF250386	AF250386 Carsonell
22	77	4.6	256172	2	AC005139	AC005139 Plasmodiu
23	77	4.6	310779	2	AC005140	AC005140 Plasmodiu
24	76.4	4.6	713	3	AF034389	AF034389 Plasmodiu
25	76.2	4.6	14867	3	AE001398	AE001398 Plasmodiu
26	76.2	4.6	224448	2	PFMALIP4	AL034477 Plasmodiu
27	75.8	4.6	108908	3	PFMALIP8	AL034560 Plasmodiu
28	75.6	4.5	12029	3	AE001431	AE001431 Plasmodiu
29	75.6	4.5	234112	3	PFMALIP2	AL034475 Plasmodiu
30	75.4	4.5	160891	9	AL512410	AL512410 Human DNA
31	75.2	4.5	86827	3	PFMALIP5	AL034556 Plasmodiu
32	75	4.5	1156	3	MPU15678	U15678 Myrmecia pl
33	74.8	4.5	222512	9	AC006335	AC006335 Homo sapi
34	74.6	4.5	186431	2	AC022281	AC022281 Homo sapi
35	74.6	4.5	213530	2	AC090493	AC090493 Mus muscu
36	74	4.4	9000	3	TETRRTRNA	L28677 Tetrahymena
37	74	4.4	14001	3	PFMALIPB	X95276 P. falciparu
38	74	4.4	47296	3	AF160864	AF160864 Tetrahyme
39	74	4.4	66441	3	PFMALIP4	AL031747 Plasmodiu
40	74	4.4	293431	2	PFMALIP4	AL049181 Plasmodiu
41	73.8	4.4	53932	2	AC023371	AC023371 Homo sapi
42	73.8	4.4	208684	2	AC072048	AC072048 Mus muscu
43	73.6	4.4	178193	2	AC068949	AC068949 Homo sapi
44	73.6	4.4	186431	2	AC022281	AC022281 Homo sapi
45	73.2	4.4	130349	9	AC011593	AC011593 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS BBU62900  
DEFINITION Borrelia burgdorferi flagellar filament outsheath protein (flaA) gene, complete cds, and chemotaxis histidine kinase (cheA) gene, partial cds.  
ACCESSION U62900  
VERSION U62900.1 GI:1575445  
KEYWORDS  
SOURCE  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.  
REFERENCE  
1 (bases 1 to 1655)  
Ge,Y. and Charon,N.W.  
An unexpected flaA homolog is present and expressed in Borrelia burgdorferi  
JOURNAL J. Bacteriol. 179 (2), 552-556 (1997)  
MEDLINE 97144545  
REFERENCE 2 (bases 1 to 1655)

AUTHORS Ge.Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1996) Yigong Ge, West Virginia University,  
Microbiology, HSCN, Morgantown, WV 26506, USA  
FEATURES Location/Qualifiers  
source 1.1655  
/organism="Borrelia burgdorferi"  
/strain="212"  
/db\_xref="taxon:139"  
/clone="pml and pm2"  
1.395  
/function="unknown"  
/note="OrfA"  
/codon\_start=3  
/transl\_table=11  
/protein\_id="A0C44769.1"  
/db\_xref="GI:1575446"  
/translation="DNLFKAVFYHSSKGFYANIRYLREDEONLKDNLVLPKDFLLG  
SNEGFSFGLSDSRFLYSFLKNGVYVNLREFYDFRNNGDNEFPVKVLFAMSL  
IKTRMFNPGKIKIILIVEGILKEOS"  
473..1498  
/gene="flaA"  
473..1498  
/gene="flaA"  
/function="putative flagellar filament outsheath protein"  
/codon\_start=1  
/transl\_table=11  
/product="FlaA protein"  
/protein\_id="A0C44770.1"  
/db\_xref="GI:1575447"  
/translation="MKRRAKSILFFLLSTVLFAQFTDGLAEGSKRAPEGLVLDFAEL  
AADPSTRDLNVDVYVSGASGVIVKEDVDVGINNWSYLTPSARLOAVYKNSV  
VAPAVKESKRYAGDTILGVLEFVPSYSSOSAMIMPEKIPFYSGESGNOLGKLI  
DNIKMKKEIKVSYLGYEIDLEVFEDMNGMEVYSGMTLKGWADLIMSNPVYIP  
NISRITKDDVNPVPLASKMKRFKFRSKSSSVYKNPIFYKRLVRLYDKLSID  
SDIDESVPKVIETSGTBSLKRKAHEPKRVLKREKISTAESSGFONFVEKTESKP  
EESSPRN"  
1566..1655  
/gene="cheA"  
1566..>1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon\_start=1  
/transl\_table=11  
/product="CheA protein"  
/protein\_id="A0C44771.1"  
/db\_xref="GI:1575448"  
/translation="MEILDLENEELGVFFEBANQMLVDILENT"  
BASE COUNT 560 a 182 c 308 g 605 t  
ORIGIN  
Query Match 98.4%; Score 1637; DB 1; Length 1655;  
Best Local Similarity 99.5%; Pred. No. 3.5e-210;  
Matches 1655; Conservative 0; Mismatches 0; Indels 8; Gaps 1;  
QY 1 atgataatcttttcaaaaaggttttattatcattcattcaaggaagattgtgcta 60  
DB 1 atgataatcttttcaaaaaggttttattatcattcattcaaggaagattgtgcta 60  
QY 61 attaagataattgaagaatgaacaatttgaagaataattagaacctttagtaaaag 120  
DB 61 attaagataattgaagaatgaacaatttgaagaataattagaacctttagtaaaag 120  
QY 121 attctctttaaggaagaatgaggtcttctcttggttttattaaagtatcaagat 180  
DB 121 attctctttaaggaagaatgaggtcttctcttggttttattaaagtatcaagat 180  
QY 181 ttttataatctttttaagaagaatgaggtttattatgtaaatctttaagaagaattatg 240  
DB 181 ttttataatctttttaagaagaatgaggtttattatgtaaatctttaagaagaattatg 240  
QY 241 atctcttaataatggtgattataatgaatctttgatgttaagcaatcttttgcta 300

DB 241 attctcttttaataatggtgattataatgaatcttttgatgattaaagctaatctttttgcta 300  
QY 301 tctctttaataaacaatggcctttaaactatcccgtaagaataaagaattatattc 360  
DB 301 tctctttaataaacaatggcctttaaactatcccgtaagaataaagaattatattc 360  
QY 361 ttgttgaagggtgtaactttaaaggacaagttgaataatlaaatttactaataaata 420  
DB 361 ttgttgaagggtgtaactttaaaggacaagttgaataatlaaatttactaataaata 420  
QY 421 aattaaaaaaccgaaatttataaagaattatataagaaggttggtttacatgaag 480  
DB 421 aattaaaaaaccgaaatttataaagaattatataagaaggttggtttacatgaag 480  
QY 481 gaaactaaagaattttatt 540  
DB 481 gaaactaaagaattttatt 540  
QY 541 tggattagcagaagggttcttaaaaggcagagccttgagaataatagcttagatttcgca 600  
DB 541 tggattagcagaagggttcttaaaaggcagagccttgagaataatagcttagatttcgca 600  
QY 601 gcttgcaagaagatccaagtccaactagacttgatcttacaattatgttattatgata 660  
DB 601 gcttgcaagaagatccaagtccaactagacttgatcttacaattatgttattatgata 660  
QY 661 ttcggagcgcttcttgattgtttaagccgagaataatgcttgtagatcttgagataataa 720  
DB 661 ttcggagcgcttcttgattgtttaagccgagaataatgcttgtagatcttgagataataa 720  
QY 721 ttgagacgttttacttactctcttcgcaaggttcgagagcttaacggttaaaattcagttg 780  
DB 721 ttgagacgttttacttactctcttcgcaaggttcgagagcttaacggttaaaattcagttg 780  
QY 781 tgcgcgcgctgttcttaagagtgagtcgaaagaatgacgaggttgatctatttgggggt 840  
DB 781 tgcgcgcgctgttcttaagagtgagtcgaaagaatgacgaggttgatctatttgggggt 840  
QY 841 aagagtttggttccaagatcttccaatcatccttcctatgattatgacacattaaat 900  
DB 841 aagagtttggttccaagatcttccaatcatccttcctatgattatgacacattaaat 900  
QY 901 tcccttttaacaggggaagtggaatcaattttaggcaaaaggttctatgtatgaatcat 960  
DB 901 tcccttttttttaggggaagtggaatcaattttaggcaaaaggttctatgtatgaatcat 960  
QY 961 taaaacatgaagaagaattaaaggtatctgttcttaagtttaggtatgagatagcttga 1020  
DB 961 taaaacatgaagaagaattaaaggtatctgttcttaagtttaggtatgagatagcttga 1020  
QY 1021 ggtttatttgaagaatgaatgagatgcaatgaaatgcttatctatgagttacttaaaat 1080  
DB 1021 ggtttatttgaagaatgaatgagatgcaatgaaatgcttatctatgagttacttaaaat 1080  
QY 1081 taaaggttgagctgatttaatttggtccaatccctaactatattccataatacatccag 1140  
DB 1081 taaaggttgagctgatttaatttggtccaatccctaactatattccataatacatccag 1140  
QY 1141 aattttaaagaagatgttccaattatcctcttgcttcaagtaaaatgagattaaagc 1200  
DB 1141 aattttaaagaagatgttccaattatcctcttgcttcaagtaaaatgagattaaagc 1200  
QY 1201 ttttagagtttcaaaagtcacagagttcaaaagttaaaatttcatcttttattgtaaga 1260  
DB 1201 ttttagagtttcaaaagtcacagagttcaaaagttaaaatttcatcttttattgtaaga 1260  
QY 1261 ttttagagtttcaaaagtcacagagtttcaaaagttaaaatttcatcttttattgtaaga 1320  
DB 1261 ttttagagtttcaaaagtcacagagtttcaaaagttaaaatttcatcttttattgtaaga 1320  
QY 1321 tgaatttaaaagttttagagctagcgaactgaatcccttcgtaaaatgaagcacaga 1380  
DB 1321 tgaatttaaaagttttagagctagcgaactgaatcccttcgtaaaatgaagcacaga 1380



Db 1321 TGTATTTAAAGTTATGAGACTAGCGGAAGTGAATCCCTTGTAAATTAAGGACACAGA 1380  
Qy 1381 aactttaaagaagctttaagaagcttaagaagaaatctcatcgctgaagctctcca 1440  
Db 1381 AACTTTAAAGAGATTTTAAAGCTTAGAGAAAAATTTCTATCGCTGAGGCTCTTCCA 1440  
Qy 1441 aaacttctgaagaagcttgaagctgaagaagaaatcatctccgaagaataggt 1500  
Db 1441 AACTCTTGTAGAAAGATTGAGAGTGAAGAACTGAAGAACATCTCCGAAAAATTAAGT 1500  
Qy 1501 ttaaatataatgttaagaagcttaagaagcttgccttctterroyasnacataltaaaata 1560  
Db 1501 TTAATTAATTAATGTAAGAGCTACCTAAAAAGCTTCTTT-----ACATATTAAATAA 1552  
Qy 1561 taggaagaatagataggaagaatataagcttggaaagaagctttagagaagcttttt 1620  
Db 1553 TAGGAAATAGATATGTGAAATATTGATTTGCAAAATGAAAGAGCTTTTAGAGATTTT 1612  
Qy 1621 tgaagaagctcaaaatcttctgaatataatcttgaagaagaatatt 1663  
Db 1613 TGAAGAAAGCTCAAAATCTTGTAGATATCTTGAAGAGATATT 1655

RESULT 2  
AE001168  
LOCUS AE001168 13271 bp DNA BCT 15-DEC-1997  
DEFINITION Borrelia burgdorferi (section 54 of 70) of the complete genome.  
ACCESSION AE001168  
VERSION AE001168.1 GI:2688598  
KEYWORDS  
SOURCE  
ORGANISM  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.  
1 (bases 1 to 13271)  
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R., Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K., Gilm,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.D., Weidman,J., Uteerback,T., Wathey,L., McDonald,L., Atлич,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi  
Nature 390 (6660), 580-586 (1997)

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

2 (bases 1 to 13271)  
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A., Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K., Gilm,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.D., Weidman,J., Uteerback,T., Wathey,L., McDonald,L., Atлич,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
Direct Submission  
Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 13271  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
66. 1100  
/gene="BB0668"  
66. 1100  
/gene="BB0668"  
/note="similar to GP:1575447 percent identity: 98.83; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="flagellar filament outer layer protein (flaa)"  
/protein\_id="AAC67025.1"  
/db\_xref="GI:2688608"  
/translation="WYMKRKAISIFLLSTVLPQETDGLAGSKRAPEGLVDF

AEIADPSSRLDLTNVDYVYGASGIVPEDMVVDLGINNMSVLLTPSARLOAVYK  
NSYAPAVYKSESKRAGDTILCYRVLPSPXSOSAMAIMPFKIPSPGSGGNOFLK  
GLIDNITKMEIKVSVISLGETIDLEFEDMGMEIAYISMGTLKRGADILSNPN  
YIPNISRIIKIDVDPNYPNPLASMKRFAPFVSKSHSEKQNFIFYVLDLQKLSV  
SIDSIDSESVFVYETSGTESIRKLKAHETFRVLKREKISMPEGSPONFEKIES  
EKPEESSPN"  
1168. 3762  
/gene="BB0669"  
1168. 3762  
/gene="BB0669"  
/note="similar to PID:1113815 SP:Q44737 GB:AE000783 percent identity: 99.31; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis histidine kinase (chea-2)"  
/protein\_id="AAC67024.1"  
/db\_xref="GI:2688607"  
/translation="METIDLENEELGVPEPEAONLVILLENIMSLDDPNNSPTID  
EIRRAHTLKSSASLDMLSDPTHEVEVFPDAIRGKXNNINDLVLLSLSDYIK  
EMIALRIDGVVYLNDISDLKSLQFLVIDDQFFIKFSDNSIKNNFCLSESDLEIR  
EGLIGQGVKLVISVFNSSNSSEVENSGKIFNLKLVGLVLTIPYEQIIEDEFLK  
RVDYLYISDIEGVKSLSDNLNLESLVDEDFVYKEDEIKRLADEIKDVLDSNFPK  
DNDFDEISDILLVEYNOKLFKVRIDPEKDPNPMATISGLMQLALSLGKIFKSI  
DSSELADKRFEDFVYLYISNTSEESIARKINLPDVSHEIKVNLSELSKSVLAKED  
DEAPFKKNKIKKNSPISVULIRDSKIDYILNVSANYSKSYNOINSEMTLPY  
NENFYDQSFQNFILIDIKIVKDGTLFLEDEISHINSLSFKERKALKDSEIV  
NSEFRLQNFKMTSGRLSRIITDHESVLTFRMLPINSIFRTRVARDLSKLTSE  
NLKMEGETELDSVIDLVDPMLHCVRNSMDQLEVEERVRKSGKACGTIILRAN  
EGNVISEIEDDGIQDPKVIKRLKGTIKEDAIYSEFELINJAPAFSTAVQYT  
DLGRGVGLVVKRSKTEKNGTLTVESEIGTIFKILPLTLVITIOGLVGSSEYI  
VPLNVLVETFRTHEDIKLENHEVYNNRDEYISVLRDKLFNTRDLSLIEFKLE  
VNTSNKIAIVVDSILGEEDEYVVKPIKDKFSSAGIVGATTLGNGVVLIDFKLE  
DLQDKTE"  
3769. 5169  
/gene="BB0670"  
3769. 5169  
/gene="BB0670"  
/note="similar to GP:1881576 percent identity: 98.93; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="purine-binding chemotaxis protein (cheW-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="GI:2688606"  
/translation="MQIKETVFGPKIIDKNSKLTNDFKVVSEFGSDHYLVDM  
OVRKIRKSNFTYVPAKKYVAGIDNIRGEIPIIDIRIMNLEFNKDIEDLVLMK  
ENDLIGIVTKINNVSIDSLSLDDPVLVSDSLINIVIGVYDNKILILNVPI  
FNNGEERKLKDPONPEKSDPVSQNDLILERSKNDPKNKYSKKDKGNNSTLNN  
TAAPNLENIRKNLKTSFNASLVNDVLEKGVKFNVDYDLPDSFLNEPFSKSG  
NLNGADCLEEFKNEIVSRSLSPNNNLISFNVLSEIGSGKETMALNALSEYVKEP  
KLVAINDLSKVETSRVSESEIGISEIYSRNSFQSPGVYKFKSEIISNVLFEYS  
DAFESDLPDLNGVLFKDCVFLDSKQDILINIIASKAIKAGALLVGDNEELKNNDY  
FIKESKRYFNLRKV"  
5181. 5666  
/gene="BB0671"  
5181. 5666  
/gene="BB0671"  
/note="similar to GP:1408273 percent identity: 98.76; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (chex)"  
/protein\_id="AAC67022.1"  
/db\_xref="GI:2688605"  
/translation="MRDYTEPLDAASSVLRMLLVENIMKPGKLSINOKIKGS  
VYIGLACVSGSIIDMDITFALFVASKLNFEEEDDDDETEMAVATLTREGNITA  
GNFVTLHAKGFVFDITPPAFITGENMKISNKGSEALVPSLPDGAITLVNIAIRER  
V"  
5706. 6146  
/gene="BB0672"  
5706. 6146  
/gene="BB0672"  
/gene="BB0672"

/note="similar to GP:1408274 percent identity: 97.95;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chemotaxis response regulator (cheY-3)"  
 /protein\_id="AACG7020.1"  
 /db\_xref="GI:2688604"  
 /translation="MIQKTTIADSSSKRGINYPDTGIPFNVLIVDSYFVKKLTQIT  
 VTSGEFNPIDPAADGEAEVAKRYNHYNPIDVITLDTPMKMGITGLCSINIEFKNR  
 VIMISALGKIDLVKDCILKAKRIFVKKPLDRAKVLQIRVMSYFK"  
 /complement(6187.6702)  
 /gene="BB0673"  
 /complement(6187.6702)  
 /note="similar to GP:1408275 percent identity: 97.56;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AACG7020.1"  
 /db\_xref="GI:2688603"  
 /translation="MKIQLIIMLALLDPEPLNARLIDISTEKRADEIKRYSYNLILIL  
 KEKYETNPSEIKSEIKLETFKFSYIMLNTQNSILNSKYNKANKYLIQSELDIKK  
 ELAKIKIRKININGIFSHSLIYTKKGFKLELYIENNAEPKIFNLNITYFLKLNQK  
 ISNEMIFPPR"  
 /complement(6699.7745)  
 /gene="BB0674"  
 /complement(6699.7745)  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="B. burgdorferi predicted coding region BB0674"  
 /protein\_id="AACG7027.1"  
 /db\_xref="GI:2688610"  
 /translation="MAKRSIFLKAIIIFISFELIEBELSTILEPIPYIRALIFLGF  
 LDFPIFIFLPLKTYKVALSQRLEIVYRNNEFDDIHLCLIPAFYSYQLNNIIVAHE  
 TLIDPMIMSLFELKFLFLNLDILEIYVNSKEKMLILIAAFRFSMSLIPPTPT  
 ITSSKIVNSIPEKQEDNIIISINAKAYIEKPKPLILIKEDODIYFYSSEIYF  
 YSSSEKRVETKQEDIDKYIQRKSDSLIGLPLETTPASPTPIFLMNFYFPAFSL  
 NPILIMKRIQDPLEYKIDPIPTLSSEKAYIELANSFNMLLKEKLNKRKRKIPL  
 EKVKIILKNQEK"  
 /complement(7739.8602)  
 /gene="BB0675"  
 /complement(7739.8602)  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="B. burgdorferi predicted coding region BB0675"  
 /protein\_id="AACG7026.1"  
 /db\_xref="GI:2688609"  
 /translation="MNTKIFPEPLVLYVLFNHSIKSLAIEKEIILNLEKNKTTA  
 TFIENDESEIKELDILVKKRKTDMITTPVLSGEIAIDENIYKPLFSEKLEKNLNLIL  
 FNPFAVSSLRKEIVYFNINSIGVITPINKLFEGKNSYIYLIQKNKALYSDDIIL  
 KNYKILKILKVIKIVKCEFGKIELFTTKNIESTSAISNDTDFLODEKTAKLISPFK  
 YETLQMSPISTLIALFASARAPFKTKNNPVAGFIGYDESMYSIKQSSRREYDRIIK  
 ELSEIAKYMK"  
 /complement(8599.9261)  
 /gene="BB0676"  
 /complement(8599.9261)  
 /note="similar to GP:142023 SP:94755 PID:1004013  
 PID:1222502 PID:1204815 percent identity: 25.58;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphoglycolate phosphatase (gph)"  
 /protein\_id="AACG7019.1"  
 /db\_xref="GI:2688608"  
 /translation="MKKACQIFMDGTIVNSIMDIAFSNSALSNLGYSTIELSKFNA

Query Match	Best Local Similarity	98.5%	0	Mismatches	9	Indels	10	Gaps	2
Matches 1246;	Conservative	Pre	12.6e-13						
401	taaatcttaccctaataaataataataaaacagaataatttataaagaatt--tatata	458							
Db	1	TAAATTTACTATATAAATAATTATAAAGCAAAATTTTATAAAGATTATATATATA	60						
Qy	459	aggagcttggttacatgaagaagaaagcctaaagatttcatcttcttattatccact	518						
Db	61	AGGAGCTTGTTTACATGAAAGAAAGCAAGTAAAGATTATTTATTTTATATCCACTG	120						
Qy	519	ttctctttgtctcaagaagactgtagttagcagaaggtgtctcaaaagggcagaag	578						
Db	121	TTCTTTTGTGCTCAAGAGACTGATGATTAAGCAGAGGTTCTAAAGGCGCAGAGCTTGAG	180						
Qy	579	aattgctcttagatttgcgagcctltgaagaagatccaaagtcttaacttagacttga	638						
Db	181	AATTTGTTTAAATTTTGCCGAGCTTGCAAGAGATCCAAAGTTCAACTGACTTGATCTTA	240						
Qy	639	caaatatggttaattatgtatatactcgggcgccttcggtatgttgaagccggaaga	698						
Db	241	CAAAATTATGTTGATATATGATATTTGGGGCGCTTCGGATATGTATAGCGCGGAAGATAGG	300						
Qy	699	ttgttagatccttgggaataaataatggagcgttttaacttaactcctctcgaaggttga	758						
Db	301	TTGTGATCTTGGATGATAATAATTTGGAGCGTTTACTTACTCCCTTCGCAAGGTTGCAGG	360						
Qy	759	cttaagcttaaaaaatccagttgtgtgcgccgcgtgtgttaagagtgagtcacaaaagtacg	818						
Db	361	CTTACGTTAAAAAATTCAGTGTGTGGCCCGCGTGTGTTAAGGTGAATGCAAAAAGTACG	420						
Qy	819	caaggtgatacattcttgggggaagaagtttbtctccaagcattctccaatcatctgta	878						
Db	421	CAGGTGATACATTTTATAGGGGTMAAGGTTTGTTCACAGCATTTCTCATCATCTGGCTA	480						
Qy	879	tgatatacgccaccatttaaaattccctcttcttattcaggggaagtgagcaatcaattttag	938						
Db	481	TGATTTATGCGACCATTTTAAAAATTCCTTTTATTTCAGGGGAAAGTGCACATCAATTTTAG	540						
Qy	939	gcaaaaggtctctatgataacattaaaccatgaaagaaatgaagtatctgttataagtt	998						
Db	541	GCAAAAGGCTTATATATTAATTAACATTTAAACCATGCAAAAGAAATTAAGGTATCTGTTATAGTT	600						
Qy	999	tagggtatgagaatagatccttgggttcttatttgaagatcatgaatgcaatgaatgctt	1058						
Db	601	TAGGGTATGAGATACATCTTGAAGTTTATTTGAAGATATGATGATGCAATGCAATATGCTT	660						
Qy	1059	attcctatgggtactttaaagtttaaagggtgggctgatttaatttbtgtccaatccctaact	1118						
Db	661	ATTCTATGGGTACTTTAAAGTTTAAAGGTGGGGCTGATTTATTTTGGTCAAAATCCTACTACT	720						
Qy	1119	ataatctctaataatcatccagaatattataaagaagatgttccaaatatatcctcttgtct	1178						
Db	721	AATATTCCTAATAATATCAATCCAAATTAATTTAAAGAGATGTTCCAAATATATCCTGTGCTT	780						
Qy	1179	caagtaaaaaatgagatttaagcctttagagtttcaaaagttcaacaagttccaaagttaaag	1238						
Db	781	CAGATTAATAATGAGATTTTAAGGCTTTTAAAGATTTCAAAGTCACACAGCTTCAAAAGAGCAAA	840						
Qy	1239	attcaatccttatagtttaaagatttaagaagttcttattgataaagctlaagtglttcaatag	1298						
Db	841	ATTTCATCTCTTTTATGTTTAAAGATTTTAAGAGTTCTTTATGATAAGTTGAGTGTTCAATAG	900						
Qy	1299	attctgatatatgacagtgagctctgattttaaagtttatgagacttagcggaactgataccc	1358						

Db 901 ATTCTGATATGACAGTGAAGCTGATTTAAAGTTATGAGACTAGCGAAGTAATCC 960  
QY 1359 ttccgtaataataagcagccgaacttttaaaagagtttaagcttagaagaaattt 1418  
Db 961 TTCGTAAATTAAGGCACGCAAACTTTAAAGCTTTAAAGCTTAAGAAATAATTT 1020  
QY 1419 ccatcgctaaagcctcttccaaacttctgtagaagaagttgagttgaaaacccgaag 1478  
Db 1021 CTAATGCTTAAGAGCTCTTCCAAACTTTGTAGAAAAGATTGAGACTGAAAACCTGAG 1080  
QY 1479 aatcatctccgaataatagtttaataatagtaagaactaacctaaaggttgcctt 1538  
Db 1081 AATCATCTCCGAAAAATAGTTTAATTAATGTAAGACTACCTAAAGCTTTCCTTT 1140  
QY 1539 errorssmacatataataataggaatagatagtagaataatagatttgaaaag 1598  
Db 1141 -----ACATATTTAAATAATAGGAAATAGATATGGAATAATTTGATTGGAATAAG 1192  
QY 1599 aagagcttttagagatttttttgaagaagctcaaatcttctgtagatatacctgagaaga 1658  
Db 1193 AAGAGCTTTTAGAGATTTTTGTGAAGAGCTCAAAATCTGTAGATATCTTGAAGAGA 1252  
QY 1659 atact 1663  
Db 1253 ATATT 1257

RESULT 3  
AE001167 10016 bp DNA BCT 15-DEC-1997  
LOCUS Borrelia burgdorferi (section 53 of 70) of the complete genome.  
DEFINITION AE001167 AE000783  
ACCESSION AE001167.1 GI:268585  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Lyme disease spirochete.  
Borrelia burgdorferi  
Bacteria: Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
1 (bases 1 to 10016)  
Fraser,C.M., Casjens,S., Huang,M.M., Sutton,G.G., Clayton,R.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D.,  
Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S.,  
Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.,  
Venter,J.C., et al.  
Genomic sequence of a Lyme disease spirochaete, Borrelia  
burgdorferi  
Nature 390 (6660), 580-586 (1997)

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
Genomic sequence of a Lyme disease spirochaete, Borrelia  
burgdorferi  
Submitted (12-Dec-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA

FEATURES  
source  
1..10016  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
gene  
84..1217  
/gene="BB0656"  
CDS  
84..1217  
/gene="BB0656"  
/note="similar to SP:P54304 PID:1303805 GB:AL009126  
percent identity: 25.62; identified by sequence  
similarity: putative"  
/codon\_start=1

gene  
CDS  
/transl\_table=11  
/product="oxygen-independent coproporphyrinogen III  
oxidase, putative"  
/protein\_id="AAC67009.1"  
/db\_xref="GI:2688591"  
/translation="MRVLLPLVLESLYINLSFCCNDFIFNRIIEELKCHLILGHP  
IKITLYIKHVDCLSRONLKFETLSKXINMLLEEETLEIIPYVFEKFLKDE  
RCITRIINLWDFCSLEKRIKINGIPEISYKLLIINIKRKPPEDINDVTMMPLKDE  
SHLRDLKELSLMEPHICFSFICEBEFVLRDPDNSIDSEKLNCAECLESNGYI  
NYETNFALKGHESRHNLKMLWELKPLGLAVSLFCGDNKNVNRALIRITGSFVK  
ANNHLVTEFLKDLDEFVYHFIQGLGTIGVSLRALRLEFVNEKQFQFINVCSTLS  
KKEFVDNIMLKGRERKLNFLYLVIIINHFDNPFVKLRIP"  
complement(1204..1890)  
/gene="BB0657"  
complement(1204..1890)  
/gene="BB0657"  
/note="similar to GB:L77117 PID:1500500 percent identity:  
39.53; identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribose 5-phosphate isomerase (rpl)"  
/protein\_id="AAC67008.1"  
/db\_xref="GI:2688590"  
/translation="MENOKILVAKHAIDHYIKSNMNLGIGTGTIYVIAIKYLSKIKS  
GSLNKLKFTTSDPRKYLKSKQIDPEVSNFLNKRLDIAIDGADLEILKSLKGM  
GSAHLMKXIAYANSETLITADETKYKLGTRKMPPIEVAQNAVGFITRLLENMLE  
ATLRICKEKKGPTITDNNVYILDKVKNHVNPEGTETKFLPLGLEIGLFINHKNRIY  
VYQDKQIKEA"  
2024..2785  
/gene="BB0658"  
2024..2785  
/gene="BB0658"  
/note="similar to SP:P31217 GB:U00096 PID:1651340  
PID:1786970 percent identity: 59.76; identified by  
sequence similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoglycerate mutase (gpmA)"  
/protein\_id="AAC67007.1"  
/db\_xref="GI:2688589"  
/translation="MLGDPMYKLVLRGSEEMKNENLFTGWTDLVLSKGDIEAVEA  
GILLKQEGSPIAPESLISLRANDITLIRLGLSGSYVKKTRMLNENHYALQGLN  
KSETAAKYEDKVLIRRSYDVPNPLSDSDDRHKPKDRYKHIPRELSPTECLKDT  
VARVLPYMTDETAKEVLEBKRVIAAHGNSLRALVTFNLSBEDVLKLNIPGLPVLV  
VELDKDLPDKIKHYLGEDESKIRKMSVASQGLK"  
complement(2853..4418)  
/gene="BB0659"  
complement(2853..4418)  
/gene="BB0659"  
complement(2853..4418)  
/note="similar to GB:L77117 SP:Q57959 PID:1591243 percent  
identity: 33.14; identified by sequence similarity:  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="lysyl-tRNA synthetase"  
/protein\_id="AAC67006.1"  
/db\_xref="GI:2688588"  
/translation="MKTAMADFYAEKIRKGGPKNLVYVAGSTPSTGVHIGNPREV  
ISVDLVARLRDGSKVRFTYSDWNDRPRKVRKNNPEDELITTYRQAITRVPRRS  
HRTSYARANIEFEKYLPAVGAIINPEIDFOSKPTVSAVSAQIRFALDHKSEALNE  
YRSTKLENNWYIPVCTCKNDTITVNNYDNHYSEYCEGNGOSLDIRTMAIKL  
PWRIDPMKRYEKVDREPAGKDHSSGGSFDSKIVIFQSGPVTYQYDITSIK  
RGKGLSSSGDVIISLKDLEVYTPETRLFAATKNTSEFSISFDLVIRIYEDYKF  
ERIVGVEDVKEKRAFRIVELSQVPMPSKRIPOVGFRIHLSVTSQLENNINIKL  
NYLKNVQEDQDKLINKINCAINMIDFAPDEKYSVLAQSPFMELIEENSKAIINEL  
LDPLKNEFVATEODIQNEIYKISRENNIPEALFFQIYKILIDREKGPKLAFITII  
GIDREKITSKYV"  
complement(4505..5377)  
/gene="BB0660"  
complement(4505..5377)  
/gene="BB0660"  
/note="similar to GB:M4658 SP:P06616 PID:416295 PID:42767  
PID:987641 percent identity: 37.85; identified by sequence

```

gene gene
CDS CDS similarity: putative"
/codon_start=1
/transl_table=11
/product="GTP-binding protein (era)"
/protein_id="AAC67005.1"
/db_xref="GI:2688587"
/translation="MKSGFAALIGRPSNGKSTLANSICGHKISTISPIPOTRRNNIKG
ITPTDRGQIIFIDIPRGSLSKKKKNIMAKNHSISGEVLLIIIDIQKGPEENK
MLEIKNSKIFLVLNKIDLNKTIKRKITDPLFKEGLESDNTIKISAERKINTEELK
NIKYNFSEGPLYPYQEYYTDQEIINFRISETIRREKAILEEKLPSLYVDIDPLELK
KGSLEIRANIEVANESOGSIIIVGNKGKEIKSIGERRARTIAKIFETPCNLFOYLKKK
NMNKEDILIKRLIN"
5467..5838
/gene="BB0661"
5467..5838
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0661"
/protein_id="AAC67015.1"
/db_xref="GI:2688597"
/translation="MRNRFLAKVNFEVECECGEKVGLRNSVCNPCAAKIGDIECPNC
RHGVGYAFGGCCPMCHSPROELKEKFRRKEARARYSDAAKTVRLEFHPOIND
ILLYFFSSFLYLEFLTYVG"
5848..6258
/gene="BB0662"
5848..6258
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0662"
/protein_id="AAC67014.1"
/db_xref="GI:2688596"
/translation="MGITVFYLFSTIFSASFUGSSMSDKENVLKSFTFYDVDFEVERP
YARKDTGLFIATKTLKAVFNFMDDKMMSSTYFPDKLISOYAIFEVKRKFGBATLV
TPLNYMLDGDISIIVLNKNRIITLKSYSINYK"
6278..6727
/gene="BB0663"
6278..6727
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0663"
/protein_id="AAC67013.1"
/db_xref="GI:2688595"
/translation="MCKLIKRFELLVSLSMFSLPADHPDYDEKIMINVCFKVJASNE
LDRAAGVGAOKVEYGGMFLVPFRKDOLSFMMENTPLLLETIAVDSGIIKEYDEL
PYSRANVSIKYRVPALDLPRGPSFKROIKIGDAVHRCFDVNSLIVE"
complement(6736..7419)
/gene="BB0664"
complement(6736..7419)
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0664"
/protein_id="AAC67012.1"
/db_xref="GI:2688594"
/translation="WLHNHTMLIARIAMINTLTFYCMTIIIFLJCSCHKNIQYDKRIKK
FLDKRKLEYKIDSNDFTLAEPDIINNKEEVIINSRLSYNSKIRIEFIGIVAFPDIN
TPKRIEIDSLMSDSYNRRVGSWEIITHNAERGINSLAYIVKAEPAFNDPLDAIDEED
IASISIFRKIIITTNNENNIDNNNECNTNESNEOPTLKOEKTNSTKSNNELKDOJEE
EELQIRTKAO"
7462..8421
/gene="BB0665"
7462..8421

```

/gene="B80655"  
 /note="hypothetical protein; identified by Glimmer;  
 putative  
 /trans\_start=-1  
 /product="B. burgdorferi predicted coding region B80655"  
 /protein\_id="AAC67011.1"  
 /db\_xref="GI:2688593"  
 /translation="MRPLGSGYTYIOGKNNYSQKSSRFSIVLSRGIKIRELFL  
 EPLNSDFIRELISIEKQNRSSLOFISEYGLKFLILSDDLNISEKYNLAKESSC

Query Match 26.2% Score 435.8; DB 1; Length 10016;  
 Best Local Similarity 97.6% Pred. No. 8,9e-50;  
 Matches 456; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

```

/organism="Borrelia burgdorferi"
/strain="212"
/db_xref="taxon:139"
/clone_1lb="Cos1"
1. .202
/gene="orf"
<1. .202
/gene="orf"
/codon_start=2
/transl_table=11
/protein_id="CAA63001.1"
/db_xref="GI:1019754"
/db_xref="SPTREMBL:044876"
/translation="DSDDSESYFVYETSGTESLRLKAHETFKRVLKREKISTAE
GSFQNFVEKIESEKPEESSPKN"
repeat_region
205. .256
/feature="hairpin"
RBS
258. .261
/gene="chea"
270. .332
/gene="chea"
270. .>332
/gene="chea"
/function="chemotactic response"
/codon_start=1
/transl_table=11
/product="histidine kinase"
/protein_id="CAA63002.1"
/db_xref="GI:1019755"
/db_xref="SWISS-PROT:Q44737"
/translation="MEIDLLENBELGVFEFAQN"
BASE COUNT      130 a      37 c      61 g      104 t
ORIGIN

```

```

Query Match      18.9%; Score 314; DB 1; Length 332;
Best Local Similarity 97.6%; Pred. No. 2.5e-33;
Matches 332; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1297 agattctgatatgacagagtcgtatttaaaagttatgagactagcggaactgaatc 1356
|||||
DB 1 AGATTCTGATATGACAGTAGCTGTATTTAAAGTTATGACGCTAGCGGAACTGAAATC 60
|||||

QY 1357 ccttcgtaaatgaagcagacgaacactttaaaagtttaaaagcttagagagaaaaa 1416
|||||
DB 61 CCTTCGTAATTTAAAGCAGACGAACTTTTAAAGAGTTTAAAGCTTAGAGAAAAAAT 120
|||||

QY 1417 ttctatcgctgaaggtcttccaaaacttctgtagaaagatgagagtgaaaacctga 1476
|||||
DB 121 TTCTATCGCTGAAGGCTCTTCCAAAACCTTGTAGAAAAAGATTGAGAGTCAAAACCTGA 180
|||||

QY 1477 agaatcatctccgaaaaaattaggttaaatatagtaagctacctaagaagttgct 1536
|||||
DB 181 AGAATCATCTCCGAAAAAATTAGGTTAAATTAATATGCTAAAGCTTAAAGGTTTGTCT 240
|||||

QY 1537 tctatgacatataataataggaatagatcataggaatattagatttgaaaa 1596
|||||
DB 241 TT-----ACATATTAAATAATAGGAATAGTATATGGAATATTAAATTGAAAA 292
|||||

QY 1597 tgaagagctttagaggttttttttgaagaagctcaaat 1636
|||||
DB 293 TGAAGAGCTTTTAGAGGTTTGTGGAAGAACTCAAAAT 332
|||||

```

```

RESULT 5
LOCUS      BBU28962      2700 bp      DNA      BCT      16-JAN-1998
DEFINITION Borrelia burgdorferi histidine kinase (chea) gene, complete cds.
ACCESSION  U28962
VERSION     U28962.1  GI:1113814
KEYWORDS
SOURCE
ORGANISM   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia

```

```

REFERENCE
1 (bases 1 to 2700)
Trueba,G.A., Old,I.G., Saint Girons,I. and Johnson,R.C.
AUTHORS
A chea chew operon in Borrelia burgdorferi, the agent of Lyme
disease
JOURNAL
Res. Microbiol. 148 (3), 191-200 (1997)
MEDLINE
98438936
REFERENCE
2 (bases 1 to 2700)
Trueba,G.A. and Johnson,R.C.
AUTHORS
Direct Submission
TITLE
Submitted (12-JUN-1995) Gabriel A. Trueba, Microbiology, University
of Minnesota, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
JOURNAL
On Dec 12, 1995 this sequence version replaced gi:984522.
COMMENT
FEATURES
source
1. .2700
/organism="Borrelia burgdorferi"
/strain="CT-1"
/db_xref="taxon:139"
95. .2689
/gene="chea"
95. .2689
/gene="chea"
/codon_start=1
/transl_table=11
/function="involved in chemotaxis"
/product="histidine kinase"
/protein_id="AAB96835.1"
/db_xref="GI:1113815"

```

```

gene
CDS

```

```

BASE COUNT      988 a      278 c      507 g      927 t
ORIGIN

```

```

Query Match      9.9%; Score 164.4; DB 1; Length 2700;
Best Local Similarity 95.3%; Pred. No. 1.9e-13;
Matches 183; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1472 cctgaagaatcatctccgaaaaaattaggttaaatatgtagtaagctactaaaggt 1531
|||||
DB 1 CCTGAAGAATCATCTCCGAAAAAATTAGGTTTAAATTATATGTAAGCTACTTAAAGGT 60
|||||

QY 1532 ttgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1591
|||||
DB 61 TTGCTTT-----ACATATTAAATAATAGGAATAGTATATGGAATATTGATTTG 112
|||||

QY 1592 gaaatgaagagctttagaggttttttttgaagaagctcaaatctgtagatcctt 1651
|||||
DB 113 GAAATGAGAGATTTTAGAGGTTTTTTTGAAGAGCTCAAAATCTTGTGATATCTCT 172
|||||

QY 1652 gaaagaataatt 1663
|||||
DB 173 GAAGAGATATTT 184
|||||

```

```

RESULT 6
LOCUS      A24079      1140 bp      mRNA      PAT      29-MAR-1995
DEFINITION T. hyodysenteriae 43 kd sheath protein gene.

```

ACCESSION A24079  
 VERSION A24079.1 GI:904392  
 KEYWORDS  
 SOURCE Brachyspira hyodysenteriae.  
 ORGANISM Brachyspira hyodysenteriae; Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.  
 REFERENCE 1 (bases 1 to 1140)  
 AUTHORS Koopman, M. B. H. and Kusters, J. G.  
 TITLE Treponema hyodysenteriae vaccine  
 JOURNAL Patent: EP 0534526-A 1 31-MAR-1993;  
 DUPHAR INTERNATIONAL RESEARCH B.V.  
 FEATURES  
 source location/Qualifiers  
 1..1140  
 /organism="Brachyspira hyodysenteriae"  
 /strain="C5"  
 /db\_xref="taxon:159"  
 /cell\_line="CBS 513.91"  
 -35\_signal 27..32  
 /note="putative"  
 -10\_signal 50..55  
 /note="putative"  
 RBS 86..91  
 /note="putative"  
 sig\_peptide 103..159  
 CDS 103..1065  
 /codon\_start=1  
 /transl\_table=11  
 /product="endoflagellar sheath protein"  
 /protein\_id="CAA01715.1"  
 /db\_xref="GI:904393"  
 /db\_xref="SWISS-PROT:P32520"  
 /translation="MKLFVVLTSIFIAASAYGLNSTLIDFALTGADNQLQAGEGDT  
 NEVVVAENLVNDNMVWLNESARLTENRRSYVTNDSKNGAMEGKVLGVVHF  
 PLAAANSYALVKPYELEMVGADGKTYTEKGKGIHNGELKISISWYGNLYLSYF  
 VNLOMERBELKSYPMGTYFNGMROVRENREYLPNRDRVLVREPLYPRMIPSKLD  
 SLGEYRKDTKGDFITVVKDVTLEYDVVVVDEEDIDDEATWOLTKTENRKAIES  
 ARIRQARLRLDEQRIDGTAADGAAANTGAADTGAOEOAO"

BASE COUNT 390 a 152 c 233 g 365 t  
 ORIGIN

Query Match 6.5%; Score 108.6; DB 6; Length 1140;  
 Best Local Similarity 49.4%; Pred. No. 6e-06;  
 Matches 369; Conservative 0; Mismatches 369; Indels 9; Gaps 3;  
 QY 581 ttatgtcttattgttcgcagcttgcagaagatccaaagttcaactgaacttgaattcttaca 640  
 DB 136 tttatcgctgcacatgcttgcgttttaacaaactcaacttttgattgatttgcgttttaaca 195  
 QY 641 aattatgttattatgtatatttcgagcgcttcgtgattgtaagccggaagatattgctt 700  
 DB 196 ggtaattgcgtatnacttacaagctgagaaagtgatnacaatgaacttgcagttgca 255  
 QY 701 gtatagcttgggataataattggagcgcttctaactcttccttcgcaagttgcagact 760  
 DB 256 gaaaaacttttaataatgatgagtgagtgatgctgaatgaatctctgatttaacagag 315  
 QY 761 taagttaaaattcagttgttcgcccgtgttgttaagagtgagtcacaaaagttacgca 820  
 DB 316 aatcgcgaaaattcttattgcttaacgtagacagtaaaagtgtaacaaacgcttgctgggaa 375  
 QY 821 ggtatcaccattttggggtaagagtttcttcacagctatctc---aatacttgct 877  
 DB 376 gcaggttaaggttcttgcttgaagagtagactttccctttagcagcttggaaacgttatgct 435  
 QY 878 atgattatgccacacattaaatctcttattatccaggggaaagtggcaatcaatt--- 934  
 DB 436 ttatgtaaaacagatattatgaactgaatgtagcggtgtagctgatttaataatataca 495  
 QY 935 ttatggaagagcttattgataacattaaacattgaagaagaattgaattatctgtttat 994  
 DB 496 gaaggttaaaaggttattatatacattgagcgaatcaaatcttattgatttcttggtttat 555

QY 995 agttaaggatgatgatgatcttgaagtttcaatttcaagatatgaatggcatggaatat 1054  
 DB 556 gcaggttaactatttatttatttgcgttaacacttaacaaatggttgcgttaataaa 615  
 QY 1055 gcttattcattggtactttaaagtttaagggttgccgatttaatttgccaatctt 1114  
 DB 616 tctttatccaatggctgcttttacttcaacggcttgagacaaatgagatggcaaacaca 675  
 QY 1115 aactatattcccaatatalatcaccagaattatataagaagatgttccaaatctctt 1174  
 DB 676 gattacttaccatnacttgcgcacagatattagtaagaaacgtct---ttatcctaga 732  
 QY 1175 gcttcaagtaaaatgagatttaaggctttagagtttcaagtcacacagttcaagtt 1234  
 DB 733 atgattcccttctgttaaatattgatttctttagcgtttctatnagaactaaagattc 792  
 QY 1235 aaaaattcattcttattatgttaagatttaagtttcttataagataagcttaagtttca 1294  
 DB 793 ggagatttcattcattcattacgttaaacgttaacacgttgactatgacgtatgattgtgat 852  
 QY 1295 atagattctgatattgacagtagtct 1321  
 DB 853 ttggaagaagatattgcacgtatgact 879

## RESULT 7

THFLA1 1260 bp DNA BCT 01-DEC-1993  
 LOCUS T.hydroysenteriae fla1 gene.  
 DEFINITION X63006 S38343  
 ACCESSION X63006.1 GI:433523  
 VERSION  
 KEYWORDS fla1 gene; Fla1 protein; flagellar sheath; flagellin; periplasmic protein.

SOURCE Brachyspira hyodysenteriae.  
 ORGANISM Brachyspira hyodysenteriae; Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.  
 REFERENCE 1 (bases 1 to 1260)  
 AUTHORS Kusters, J. G.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-Nov-1991) J. G. Kusters, Inst. of Infectious Diseases and Immun., P.O.Box 80.165, 3508 TD Utrecht, THE NETHERLANDS

REFERENCE 2 (bases 1 to 1260)  
 AUTHORS Koopman, M. B., de Leeuw, O. S., van der Zeijst, B. M. and Kusters, J. G.  
 TITLE Cloning and DNA sequence analysis of a Serpulina (Treponema) hyodysenteriae gene encoding a periplasmic flagellar sheath protein  
 JOURNAL Infect. Immun. 60 (7), 2920-2925 (1992)

MEDLINE 92307926  
 FEATURES

## source

Location/Qualifiers  
 1..1260  
 /organism="Brachyspira hyodysenteriae"  
 /strain="C5"  
 /db\_xref="taxon:159"  
 -35\_signal 27..32  
 -10\_signal 50..55  
 RBS 86..91  
 gene 103..1065  
 /gene="fla1"  
 sig\_peptide 103..159  
 /gene="fla1"  
 CDS 103..1065  
 /gene="fla1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="flagellin Fla1"  
 /protein\_id="CAA4735.1"  
 /db\_xref="GI:433524"  
 /db\_xref="SWISS-PROT:P32520"  
 /translation="MKLFVVLTSIFIAASAYGLNSTLIDFALTGADNQLQAGEGDT  
 NEVVVAENLVNDNMVWLNESARLTENRRSYVTNDSKNGAMEGKVLGVVHF  
 PLAAANSYALVKPYELEMVGADGKTYTEKGKGIHNGELKISISWYGNLYLSYF  
 VNLOMERBELKSYPMGTYFNGMROVRENREYLPNRDRVLVREPLYPRMIPSKLD  
 SLGEYRKDTKGDFITVVKDVTLEYDVVVVDEEDIDDEATWOLTKTENRKAIES  
 ARIRQARLRLDEQRIDGTAADGAAANTGAADTGAOEOAO"

Query Match	6.5%	Score 108.6	DB 1	Length 1260
Best Local Similarity	49.4%	Pred. No. 5.9e-06		
Matches 369	Conservative 0	Mismatches 369	Indels 9	Gaps 3
<div> <div>mat_peptide</div> <div>160. .1062</div> <div>/gene="flaI"</div> </div> <div> <div>repeat_region</div> <div>1004. .1030</div> <div>/product="flagellin FlaI"</div> </div> <div> <div>repeat_region</div> <div>1157. .1175</div> <div>/rpf_type=INVERTED</div> </div> <div> <div>repeat_region</div> <div>1157. .1175</div> <div>/rpf_type=INVERTED</div> </div> <div> <div>BASE COUNT</div> <div>437 a 158 c 245 g 420 t</div> <div>ORIGIN</div> </div>				
OY 581	tttagcttaagatttggccgagcttgcaagaagatccaaagttcaactagactatcttca	640		
DB 136	TTTATCGTCGCATCTGCTTACGGTTTAACTCAACTGATGATTTGCTTTAACA	195		
OY 641	aattatgtttgatlatglatcatctcgagcgcttcgtgtatgtttaagccggaagatagtt	700		
DB 196	GGTATGCTGTAATCACTTCAACAGCTGGAAGAGTGATACAAATGAAGTACGTTGCA	255		
OY 701	gttagctcttgggataaataatttggagcgcttactactcccttcgcaagttgcagct	760		
DB 256	GAAATCTTTTATATGATGATGACTGGGTACTATGTTGTAATGTAATCTGATGATCA	315		
OY 761	tactgttaaaatctcagttgttcgcccgtgtgtttaagagtgagtcataaaggtaacga	820		
DB 316	AATCGCAGAAATTTCTTATGTTACTTACGTAACGTAGACATTAAGTAACAGCGTCTG	375		
OY 821	gtgtactatatttggggtaagagtttggttccaagctatctc--aatcatctgct	877		
DB 376	GCAGGTAAGTCTTGGGTGAAGATGATCATTTCCATTAGCAGCTTGGAACTGATGCT	435		
OY 878	atgatatgacccacttaaaatcctcttatactcaagggaaagtggaatcaattt---	934		
DB 436	TTAGTAAACACAGTATGACTTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA	495		
OY 935	tttagcaaaagtcctattgataacaataaaccatgaaagaataaaggtatcgttcat	994		
DB 496	GAAAGTAAAGGTGTATACATCAATCGTGGCAAAATCAATCATTAATGTTCTTGGGTTT	555		
OY 995	agtttaagggtatcagatagatccttgaagtttatttgaagataatgaatgcatgtaatt	1054		
DB 556	GGAGCTAACTATTATTAATTAGTTATTTCGTAACCTTAACTTAACTTAACTTAACTT	615		
OY 1055	gcttacttcatgggtacttlaaagtttlaaagggtggcgatcttattatgtgcaaatctt	1114		
DB 616	TCTTATCCAAATGGGTACTGTTACTTCAACGGTGTGAGCAGATGAAGTGGAAACAGA	675		
OY 1115	aactatattcctaaatataatcatccagaattataaagaagatgltccaaattatcctctt	1174		
DB 676	GAAATCTTAACTTAAAGTTGGCGACAGAGATTAATGTAAGAGAACCTT--TTATCTT	732		
OY 1175	gcttcaagtaaaatagaatttlaagcctttagagtttcaagttcaagtcacacagttcaaa	1234		
DB 733	ATGATCCCTTCTGTTAAATTAGATCTTTAGTTTCTATAGAACATTAAGATCTTAAAGC	792		
OY 1235	aaaaattcattccttattatgtttaagaatttaagagttcttattagataaagctaaagtttca	1294		
DB 793	GGAGATTTCATCACACTTACGTTAAGATGTAACACTTGTAGTATGAGTATGTTGTGAT	852		
OY 1295	atagttctcgtatgatgacagtgagctt	1321		
DB 853	TTTGAAGAAATGATGACGATGAACCT	879		

VERSION	KEYWORDS	SOURCE
AL031746.9	HMG.	malaria parasite P. falciparum.
GI:6594243		Plasmodium falciparum Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 67970) Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quall,M. and Barrell,B.
TITLE	JOURNAL	COMMENT
Direct Submission	Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK	On Dec 16, 1999 this sequence version replaced gi:5763807. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P.falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.
FEATURES	SOURCE	location/Qualifiers
gene	CDS	1..67970 /organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /chromosome="1" complement(join(1748..2598,2748..2848,2990..3276)) /gene="MAL1P3.01" complement(join(1748..2598,2748..2848,2990..3276)) /gene="MAL1P3.01" /note="MAL1P3.01, conserved hypothetical protein, len: 412 aa, similarity: UPF006 family eg to YB1055C/YB10512/YBL0511, YPF5_YEAST (418 aa), fasta scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa overlap)" /codon_start=1 /product="conserved hypothetical protein, UPF006 family" /protein_id="CAB63556.1" /db_xref="GI:6594244" /translation="MKLVFHYKYINLVFYISIIFLKSNLSKIYNDRISTVKKRYKV LQIKRSNLKNHNIIRKMEDNESSFIIGSLNLTDMKFQGVNSKRKHENDLOAVNRAR NNNDLIITCTCAETIDSLKICETDPDEGKFLYSAGVPTCYEFIDNKKEEK LIKKEYEEFTKYRKEOVSENMENKRIIOCEKGMNMLNELLEKNDITGPFRY NEKDREYLELNKKKIITYPNRIVYTIGEIGDPLDYCSKITIQIKRTIFOLKLYOMRN LPQLMHRNCSETFEKIVDIYKFLPEFGNGVTSHFTREDIVHLIYVONKLYIGVNG CSKLLENINAVKRKIPDLILLTADAPCGVKTHASVEYIKDYERAAVTNLRKRN LIKDDNTIFEREHPYNIA" complement(2599..2610) /note="potential splice acceptor sequence" complement(2742..2747) /note="potential splice donor sequence, atg/gtcaaa" complement(2849..2861) /note="potential splice acceptor sequence" complement(2984..2989) /note="potential splice donor sequence, aaa/gtaaaa" 5005..5496 /gene="MAL1P3.02" 5005..5496 /gene="MAL1P3.02" /note="MAL1P3.02, hypothetical protein, len: 163 aa, contains possible signal sequence" /codon_start=1 /product="hypothetical protein, MAL1P3.02" /protein_id="CAB63557.1" /db_xref="GI:6594245" /translation="MKLLNNRFVVLCPIILFFELNSVLGNRRNNINFHEENAKK AMRLSGEINSIKLDNGDELKILDNDEKHKDSYKWDKSYFSINLEEEKYSQDLPFR KKEINANTRKIIEEDREFYLINNDENIATRVLENNDDELAYIQSFQSLIDIQS LN"
misc_feature		8020..10389 /note="possible cent, region of very high [A+T] content" 14884..20352
gene		

```

CDS
    /gene="MAL1P3.03"
    /protein_id="14884..20352"
    /note="MAL1P3.03, putative ABC transporter, len: 1822 aa"
    /codon_start=1
    /product="putative ABC transporter"
    /protein_id="CAB63558.1"
    /db_xref="GI:6594246"
    /translation="MTTYKENVGISNKKKKKKSCONISFLNPLSPDMIRPLINDLIK
    GDIOELPNICRNDVDPVYASKLEPNRLDIEVDSEFSKSSNENHVLHHCNSASE
    KKVVVYVNIIMSLTKTKPKRIILISFYILETLVLTGKFRIDYIMALEGQKITP
    YLSELDKEVFSGLVAVMIMFHLFEALHLYFHLFTINLVLSMTELYKINLCSNN
    NNLQNDPAEYNTRRKFSQTEIDELSRDLISGKNASSSGIKNNKKINDKKNEVEN
    DYIINEIKSTKMKEDSLNERSLNPVNIYIMSPVPSFTVSCILFVYVAVLE
    MSFVEYHIGKNSGIAIMLSIALYSAMILEPFLSPKSKYLYIRKDIRDMHVL
    KEFKIKMMESPAKYINIRMKEMKCKIRLYINIGVFISSISSTIVAVYFPI
    YLKDRLNKEELKFTSIMPLVYKYLISNVANEPVLVANNMEGYNTKRLNINIDH
    LYNDIKNTFMYRTNEDYNIYVDTQLQNTITSHDGTSHNLKHLKNVILKNLTN
    MKYFEFYHKMNHKNIINKQILSGKLVNDNTNKKICFOEHSNSTYNSSHTHE
    KKEEYENHNSNSTMSNEFEKKKKNEYIKLENSCFGLSYDNKDNHILKNIENF
    LKNSLAIITGVNGSKSAFHSIIGDEFMTHGNLYIENFEKMPILVYVPOKSWLFG
    NIRMILTCNEYNPLIKYITLLOSELNDLSPIEHGDMKYINDHNLSTGQVRCILA
    RALYEHYHMKCLDYDEKLLQPELIDKDLINKNISYNNKSKLVNIPENEN
    YLQCLMDNNFYLYLDDIFSLDPSISKLFSLCEKEDIISFKDNCSTIISMKS
    TLQNLIEDLNDVQYEVNIPEIDQTKLRGNISEYMEKNMLNITKESMGYSMLT
    IDYTRIKLEDEVLNHHKSNMAYKEAFVGNTESEFSEIDSLKEIKKMKKNY
    KEHNNKNNKNNNNNNKDDHINNNNDHNRNNDINLPGNSDSDSPYSSIGNE
    YLDTYITSNNSDKKEELVPLKYDTHHEEFKSSMPEVSSNMNNPSPKEDNSS
    FKGSISLETYLMYPOVGVPLTSVYIPLISITFDEIKFVPLTMMSTISKNNKESD
    TILQOVRLTEFLVPLITSLVTSGICFSMIITGNTSAIKVNNHNLVLSILNAPLYIF
    YNNNGININRPIIDISADYGLKRIYKAFPIFPCILSLSLIITMIDCJFIPEPV
    ILIIFYEYKRRSGRCKEAORLYLSCHPCLPCLMYSALSGKMLINIKKNTYHLVYE
    HLYNNRISYFPMKLINIMASLYIKFILLFTYIIMPHLYASGIIKLYKKNVRI
    LSTGICISFSAFGLVITIKLIDYTHIEKMCVORLEEFKINSKEMASNNKNEML
    NITTOYKKEKENTSDKISATVEYKNNLSSTINSODESKKKGITGEVYVYSYK
    KPIPLNGYKTIIDEPRLKNNMTALKQKIGIVGKSGASTILLSLGLINISQG
    KITVEGRDIRYNRKGEDSIIGILAOSSVFVFNMMRTFIDYNNDEPCEIVHAKLN
    GINKGNLDKYMHKODMKNYKTIQTSKVINOSNDNYILLTNCIIRLISLVRILN
    RHRYKILIDELPIFNLNSVDELNSFLIGKASNNYIIRNHFPNNVTLISHANT
    LSCDYIYVLRKSEITRCSEYDVKTQSLSLLEMD"
    23896..31533
    /gene="rRNA"
    /note="region containing small subunit, 5.8S and large
    subunit rRNA genes and spacer regions"
    23896..31533
    /gene="rRNA"
    /complement(join(31966..32476,32675..32775))
    /gene="MAL1P3.04"
    /complement(join(31966..32476,32675..32775))
    /note="MAL1P3.04, conserved hypothetical membrane protein,
    len: 203 aa, similarity: P. falciparum chromosome 2,
    PF08110W, 096126 predicted integral membrane protein (255
    aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
    in 191 aa overlap)"
    /codon_start=1
    /product="conserved hypothetical membrane protein,
    MAL1P3.04"
    /protein_id="CAB63559.1"
    /db_xref="GI:6594247"
    /translation="MKKSTYFINVTLLFLTLLEPLTYNYNTFSKYKNNIKIDIN
    RPKRIIAESSEQKPYWEDFCLINEEELIPREHDSYLDPEHENDIKLDELINS
    TKIMKEITKRMQNYEKETDNMNNHMRDPMHWYKANNIYLVVHLINITLNDLNP
    HDKEETITWIMIOEDIEYFLFNLOVEARILFLELFYKKE"
    /complement(32477..32486)
    /note="potential splice acceptor sequence"
    /complement(32669..32674)
    /note="potential splice donor sequence, aag/gtataa"
    join(36657..36743,36864..37343)
    /gene="MAL1P3.05"
    join(36657..36743,36864..37343)
    /gene="MAL1P3.05"
    /note="MAL1P3.05, hypothetical protein, len: 188 aa"

```

```

misc_feature
    /note="potential splice donor sequence, aag/gtataa"
    36854..36863
misc_feature
    /note="potential splice acceptor sequence"
    /complement(join(38049..39955,40210..40284))
    /gene="garp"
    /complement(join(38049..39955,40210..40284))
    /gene="garp"
    /note="MAL1P3.06, garp, len: 673 aa, similarity: almost
    identical to GARP_PLAF (678 aa), fasta scores: 97.6%
    identity in 678 aa overlap"
    /codon_start=1
    /product="hypothetical garp protein"
    /protein_id="CAB63561.1"
    /db_xref="GI:6594249"
    /translation="NVLPLSYNICILFVYVCTNLSFCPSNGLLKQNIINKSEDS
    ITGRLLNTELEKNDKDSSTLLKEKDEKDVPTTSNDLKNANNNETISSTDP
    TMIIVNDRKDNESYDKRKKKDKKKKKDKKKKKDKKKKKKKKKKKKKKKKK
    DKKKENSEVMSLYTGOHKPKNATEHGEENLYEEMVSEINNAOGGLLSPYRE
    OGGCITISVHTESNDTQNDKNEISSEKKEHOOEMLKTLDEKROKKKEKKE
    KIEKKKKOBEKKEKKOKEKKKKOKEKKKKOKEKKKKOKEKKKKOKEKKKK
    HKKENETMOQDQTSSEETNNINWPLTSPLDVTTPPEHGEKEEHNHGEKEG
    BKREBEHNLEBKREKSKKESKSGKDKGKDKGKDKKREKVKHNVKNLTDE
    DKDVEYIINLEDEKACEOHITVESESPQCKLIDPEOJTLTKDKKVEKNLSIO
    EOLITICRVNVPRRNNHKKMAITEAELOKHVHDEKKEEKEEVEEVEEVO
    DEEVEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE
    DDADDEDDDAEEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
    /complement(39996..40005)
    /note="potential splice acceptor sequence"
    /complement(40204..40209)
    /note="potential splice donor sequence, aag/gtataa"
    join(45401..46396,46562..50233)
    /gene="MAL1P3.07"
    join(45401..46396,46562..50233)
    /gene="MAL1P3.07"
    /note="MAL1P3.07, hypothetical protein, len: 1555 aa"
    /codon_start=1
    /product="hypothetical protein, MAL1P3.07"
    /protein_id="CAB63562.1"
    /db_xref="GI:6594250"
    /translation="WEQIGIANNITTEIGKNIVKSGPFETIRKKFTIEVSKKISNSL
    NIENRKECKNCRKINWMDKDEEFVSKKEFVEYETISNVWKNFEFAIINTLNKT
    KTLRSRRKKYPERIINNGELEDYIDAINDYKQKPKNYCWSERYIDYKWLMEKKE

```

Query Match	5.6%	Score 93.2	DB 3	Length 67970
Best Local Similarity	50.0%	Pred. No. 0.00043		
Matches 260	Conservative	0	Mismatches 258	Indels 2
Gaps				1

```

QY 1 atgataatctttttcaaaaaggttttattatcttcctgaagggatttgcta 60
    || || || || || || || || || || || || || || || || || ||
DB 8406 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8347
    || || || || || || || || || || || || || || || || || ||
QY 61 attaagaatcttaagaagaacaattgaagaataattagatctttaagaagaag 120
    || || || || || || || || || || || || || || || || || ||
DB 8346 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8287
    || || || || || || || || || || || || || || || || || ||
QY 121 atttccttttaggaagaatgaggggttccttccttggtttattatgaagttaagaat 180
    || || || || || || || || || || || || || || || || || ||
DB 8286 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8227
    || || || || || || || || || || || || || || || || || ||
QY 181 ttattatcttttttaagaagaatgagttattatgtaaacctttcaagaagaatttaag 240
    || || || || || || || || || || || || || || || || || ||
DB 8226 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8167
    || || || || || || || || || || || || || || || || || ||
QY 241 attcctttaataatgtagtataatgaatctttgtagtataagtaaaccttttgcta 300

```



```

Db 8166 TTTATTATTAATATATATTTGTTAATTTATATATTTATATATTTTATAT 8107
      |||||
Qy 301 Tgctttaataaacaatgcgttaactatccgtgaagataaagaattattc 360
      |||||
Db 8106 TTTATTATTAATTAATTCATTATATATATATATATATATATATATATAT 8049
      |||||
Qy 361 tcttgaaagggtctatcctaagaagcgaattgataaatttactaataaat 420
      |||||
Db 8048 TTTCTTAATTAATTTATTAATTAATGTCATTAATTAATTAATTAATTAAT 7889
      |||||
Qy 421 aattaaaaacgaattatataaagattatataaaggagctgttacaagaa 480
      |||||
Db 7988 ATATATTAATTAATTAACATTAATAATTAATGGAATATATATATATATAT 7929
      |||||
Qy 481 gaaagctaaagatattattttttatataccactgtt 520
      |||||
Db 7928 ATTAATATATATTAATTAATTAATTAATCATCATGTTT 7889
      |||||

RESULT 9
LOCUS AC006986 173957 bp DNA PRI 14-OCT-2000
DEFINITION Homo sapiens BAC clone RP11-155J5 from Y, complete sequence.
ACCESSION AC006986
VERSION AC006986.3 GI:10801454
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 173957)
AUTHORS Sulston,J.E. and Waterston,R.
JOURNAL Toward a complete human genome sequence
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
99063792
REFERENCE 2 (bases 1 to 173957)
AUTHORS Joshu,C., Mohlman,P. and Phillips,A.
JOURNAL The sequence of Homo sapiens BAC clone RP11-155J5
          Unpublished
          3 (bases 1 to 173957)
REFERENCE 3 (bases 1 to 173957)
AUTHORS Waterston,R.H.
JOURNAL Direct Submision
          Submitted (05-MAR-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          4 (bases 1 to 173957)
REFERENCE 4 (bases 1 to 173957)
AUTHORS Waterston,R.H.
JOURNAL Direct Submision
          Submitted (05-MAY-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          5 (bases 1 to 173957)
REFERENCE 5 (bases 1 to 173957)
AUTHORS Waterston,R.
JOURNAL Direct Submision
          Submitted (22-OCT-1999) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          6 (bases 1 to 173957)
REFERENCE 6 (bases 1 to 173957)
AUTHORS Waterston,R.
JOURNAL Direct Submision
          Submitted (21-DEC-1999) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          7 (bases 1 to 173957)
REFERENCE 7 (bases 1 to 173957)
AUTHORS Waterston,R.
JOURNAL Direct Submision
          Submitted (14-OCT-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On Oct 14, 2000 this sequence version replaced gi:4753246.
          -----Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: saplens@wustl.wustl.edu
COMMENT

```

----- Summary Statistics  
Center project name: H\_NH015505

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frenken,E., Tateo,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)  
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is RP11-160K17. Actual start of this clone is at base position 1 of RP11-155J5; actual end is at base position 173957 of RP11-155J5.

## FEATURES

```

source
Location/Qualifiers
1..173957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-155J5"
/clone_lib="RPCI-11"
/rpt_family="AT_r1ch"
1281..1435
repeat_region
/rpt_family="Alu"
1438..1562
repeat_region
/rpt_family="(TA)n"
1621..2154
repeat_region
/rpt_family="(TA)n"
2160..7755
repeat_region
/rpt_family="L1"
7758..10708
repeat_region
/rpt_family="L1"
11258..11377
repeat_region
/rpt_family="L2"
11934..12039
repeat_region
/rpt_family="(TTATATG)n"
12046..12950
repeat_region
/rpt_family="L1"
15316..15479
repeat_region
/rpt_family="(TA)n"
23465..23506
repeat_region
/rpt_family="AT_r1ch"
23633..23653
repeat_region
/rpt_family="AT_r1ch"
23961..23996
repeat_region
/rpt_family="(TG)n"

```

repeat_region	24034. .24234	/rpt_family="Alu"
repeat_region	27018. .27131	/rpt_family="Alu"
repeat_region	27132. .27154	/rpt_family="Alu"
repeat_region	27155. .27312	/rpt_family="(CA)n"
repeat_region	27324. .27617	/rpt_family="Alu"
repeat_region	29398. .29681	/rpt_family="Alu"
repeat_region	29682. .29720	/rpt_family="Alu"
repeat_region	29721. .29736	/rpt_family="(CA)n"
repeat_region	35366. .35796	/rpt_family="Alu"
repeat_region	35775. .36379	/rpt_family="Retroviral"
repeat_region	36749. .36868	/rpt_family="Retroviral"
repeat_region	37699. .37719	/rpt_family="AT_rich"
repeat_region	38833. .38860	/rpt_family="AT_rich"
repeat_region	39860. .40054	/rpt_family="MIR"
repeat_region	40141. .40187	/rpt_family="(CA)n"
repeat_region	40832. .41426	/rpt_family="L2"
repeat_region	41431. .41523	/rpt_family="Alu"
repeat_region	41555. .41587	/rpt_family="Alu"
repeat_region	41589. .44483	/rpt_family="r-rich"
repeat_region	43510. .43651	/rpt_family="L1"
repeat_region	43652. .43952	/rpt_family="Alu"
repeat_region	44279. .46197	/rpt_family="L2"
repeat_region	49198. .49332	/rpt_family="L1"
repeat_region	49333. .50487	/rpt_family="Alu"
repeat_region	51792. .53073	/rpt_family="L1"
repeat_region	52074. .52105	/rpt_family="Alu"
repeat_region	52192. .52482	/rpt_family="(CA)n"
repeat_region	52511. .52829	/rpt_family="MER1_type"
repeat_region	53366. .53426	/rpt_family="MER4-group?"
repeat_region	54604. .54834	/rpt_family="L2"
repeat_region	55195. .55566	/rpt_family="AT_rich"
repeat_region	55792. .55940	/rpt_family="L2"
repeat_region	56086. .56299	/rpt_family="L2"
STS	complement(56224. .56702)	
repeat_region	56680. .56604	/db_xref=G1:1040785"
repeat_region	57105. .57136	/rpt_family="L2"
repeat_region	57159. .57244	/rpt_family="(TG)n"

```
repeat_region /rpt_family="(TA)n"
57276.57309 /rpt_family="(TG)n"
```

Query Match	5.4%;	Score 89.4;	DB 9;	Length 173957;
Best Local Similarity	49.9%;	Pred. No. 0.0012;		
Matches 225;	Conservative 0;	Mismatches 226;	Indels 0;	Gaps 0;

Oy	60	aatttaagatctttaagaagtgcacaaacttggaaagaatacttgaccttttggtaa	119
Db	2123	AAATTAATAATTATATATTTTAAAAATAAATAATATATATATTTGTATAAAAATPA	2054
Oy	120	gatttcctttaaggaaagcaatgaagggtcttcctcttgggtcttctaagtcattcaaga	179
Db	2053	ATATATATATATTTTATATATATAAAAACAATATATATATATTTTATATATAAAACAAT	2004
Oy	180	ttttaatactcctttttaagaaatggagttcatacgtaacctctccaagagaattcat	239
Db	2003	ATATAAATATTTTTATATATAAAAACAATAATATAAATATTTTATGTAATAAAAACAAATAT	1944
Oy	240	gatctctttaataatgltgltctaatagtaccccttgtagtgaagtcaccccttgc	299
Db	1943	ATATTTTAGTATAAAAATAATATATATATATTTTATATAGTAAAAATAATATATATAT	1884
Oy	300	atgctcttaataaaaacaalgcgccttaactacatccctgtaagataaaaaagaattatc	359
Db	1883	ATTATATATATGAATAATTAATATATATATATATATATATATATATATATATATAT	1824
Oy	360	cttgttaagggtggtatccttaaaagsgcaaatgltglaattaactttaccataaaaa	419
Db	1823	TATCAAAATTAATATATTTTCATGTATTCATATATGAAAAATAATATATTTTCATGTATCA	1764
Oy	420	taacttaaaaaagaaacttttctaataaagaattcataataaaggagcttgggtttacaagaa	479
Db	1763	TATATGAAAACAATATATTTTCATGTATTCATATATGAAAAATAATATATTTTCATGTAT	1704
Oy	480	ggaaagctaaagatcttatctttttttatc	510
Db	1703	CATATATGAAAAATAATATATTTTCATGTAT	1673

RESULT 10  
PFMAL4P1.0  
MPCOMMENT  
Sequence split into 4 fragments LOCUS PFMAL4P1 Accession AL034557

Sequence split into 4 fragments				LOCUS PFMALAP1	Accession AL034557
Fragment Name	Begin	End			
PFMALAP1_0	1	110000			
PFMALAP1_1	100001	210000			
PFMALAP1_2	200001	310000			
PFMALAP1_3	300001	392633			
LOCUS PFMALAP1	392633 bp	DNA	HTG	11-AUG-1999	
DEFINITION	Plasmidium falciparum chromosome 4 strain 3D7, *** SEQUENCING IN PROGRESS ***. In unoriented pieces.				

ACCESSION AL034557  
VERSION AL034557.7 GI:5731897  
KEYWORDS HTG; HTGS\_PHASE1.

SOURCE	malaria parasite <i>P. falciparum</i> .
ORGANISM	<i>Plasmodium falciparum</i> Eukaryota; Alveolata; Apicomplexa; Haemosporida; <i>Plasmodium</i>

**REFERENCE** 1 (bases 1 to 392633)  
**AUTHORS** Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M., and Barrall, B.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium

COMMENT  
The Sanger Centre, Wellcome Trust genome Campus, Hinxton, Cambridge  
CB10 1SA, UK  
On Aug 12, 1999 this sequence replaced gi:5531346.

On Aug 12, 1999 this sequence version replaced gi:5531346. For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum). IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be



AC079314	LOCUS	AC079314	242513 bp	DNA	HTG	09-AUG-2001
AC079314	DEFINITION	Human sapiens chromosome Xp clone RP11-386120, WORKING DRAFT				
AC079314	ACCESSION	AC079314				
AC079314	VERSION	AC079314 .25	GI:15042762			
AC079314	KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
AC079314	SOURCE	human.				
AC079314	ORGANISM	Human sapiens				
AC079314	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 242513)				
AC079314	AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., All-ossan,F.R., Allen,C., Alshrocks,S.L., Amaraltinge,H.C., Are,J.R., Banks,T., Barbara,J., Belton,J., Blmage,K., Blankenburg,K., Bonnih,D., Bouck,J., Bowe,S., Brivela,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day,C., Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovy,C., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louisged,L., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Metzger,M., Miner,G., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okwou,L., Ofagunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Peterson,L., Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmahl,K., Vasquez,L., Vera,V., Villalona,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D., Weinstock,G. and Gibbs,R.				
AC079314	TITLE	Direct Submission				
AC079314	JOURNAL	Unpublished				
AC079314	AUTHORS	2 (bases 1 to 242513)				
AC079314	TITLE	Worley,K.C.				
AC079314	JOURNAL	Direct Submission				
AC079314	COMMENT	Submitted (27-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 31, 2001 this sequence version replaced gi:14971136. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: HCB Center clone name: RP11-386120 Sequencing vector: Plasmids; M7789 Sequencing vector: M13; L08821 Chemistry: Dye-Primer Body; 4% of reads Chemistry: Dye-terminator Big Dye; 96% of reads				

Assembly program: Phrap; version 0.990329		
Consensus quality: 247540 bases at least Q40		
Consensus quality: 264763 bases at least Q30		
Consensus quality: 273200 bases at least Q20		
Estimated insert size: 327546; sum-of-contigs estimation		
Quality coverage: 0x in Q20 bases; agarose-fp estimation		
Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation		
-----		
* NOTE: Estimated insert size may differ from sequence length		
* (see <a href="http://www.hgsc.com/tmc/edu/docs/genbank_drafc_data.html">http://www.hgsc.com/tmc/edu/docs/genbank_drafc_data.html</a> )		
* NOTE: This sequence may represent more than one clone.		
* NOTE: This is a 'working draft' sequence. It currently		
* consists of 28 contigs. The true order of the pieces		
* is not known and their order in this sequence record is		
* arbitrary. Gaps between the contigs are represented as		
* runs of N, but the exact sizes of the gaps are unknown.		
* This record will be updated with the finished sequence		
* as soon as it is available and the accession number will		
* be preserved.		
1	65488: contig of 65488 bp in length	
*	65489	gap of unknown length
*	65589	129085: contig of 63497 bp in length
*	129086	129185: gap of unknown length
*	129186	161617: contig of 32432 bp in length
*	161618	161717: gap of unknown length
*	161718	171830: contig of 10113 bp in length
*	171831	171830: gap of unknown length
*	171931	175095: contig of 3165 bp in length
*	175096	175195: gap of unknown length
*	175196	179205: contig of 4010 bp in length
*	179206	179305: gap of unknown length
*	179306	184288: contig of 4983 bp in length
*	184289	184388: gap of unknown length
*	184389	188283: contig of 3895 bp in length
*	188284	188383: gap of unknown length
*	188384	192040: contig of 3657 bp in length
*	192041	192140: gap of unknown length
*	192141	195141: contig of 3001 bp in length
*	195142	195241: gap of unknown length
*	195242	198355: contig of 3114 bp in length
*	198356	198455: gap of unknown length
*	198456	201722: contig of 3267 bp in length
*	201723	201822: gap of unknown length
*	201823	205001: contig of 3179 bp in length
*	205002	205101: gap of unknown length
*	205102	207118: contig of 2017 bp in length
*	207119	207218: gap of unknown length
*	207219	209658: contig of 2640 bp in length
*	209659	209958: gap of unknown length
*	209959	212786: contig of 2828 bp in length
*	212787	212886: gap of unknown length
*	212887	215444: contig of 2558 bp in length
*	215445	215544: gap of unknown length
*	215545	217579: contig of 2035 bp in length
*	217580	217679: gap of unknown length
*	217680	220431: contig of 2752 bp in length
*	220432	220531: gap of unknown length
*	220532	222696: contig of 2165 bp in length
*	222697	225582: gap of unknown length
*	225582	225582: contig of 2786 bp in length
*	225583	225682: gap of unknown length
*	225683	227743: contig of 2061 bp in length
*	227744	227843: gap of unknown length
*	227844	230046: contig of 2203 bp in length
*	230047	230146: gap of unknown length
*	230147	232313: contig of 2167 bp in length
*	232314	232413: gap of unknown length
*	232414	234578: contig of 2165 bp in length
*	234579	234678: gap of unknown length
*	234679	237292: contig of 2614 bp in length
*	237293	237292: gap of unknown length
*	237988	239988: contig of 2596 bp in length
*	239989	240088: gap of unknown length

```

FEATURES      * 240089 242513: contig of 2425 bp in length.
SOURCE        Location/Qualifiers
              1..242513
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="Xp"
               /clone="RP11-386120"
BASE COUNT    71180 a 49220 c 48511 g 70846 t 2756 others
ORIGIN

Query Match      4.9%; Score 81.2; DB 2; Length 242513;
Best Local Similarity 49.0%; Pred. No. 0.015;
Matches 220; Conservative 0; Mismatches 228; Indels 1; Gaps 1;

Qy 50 attgtgtcatttaagatttaagagatgcacaaatttgaagataattagatct 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128435 ATATATATATAATTTNANNAGATTNTATANAATAATTAATAATNTTATA 128494

Qy 110 tttagtaaaagatttcttcttggaagcaatgaggggttcttcttggttttataag 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128495 TTTTATATATAATTAATAATTAATAATTAATAATTTTATNTATTAATTAAT 128554

Qy 170 tgaatcaagatttcttcttctttaaagaatgagatttataatccttcaag 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128555 TTATTAATATATATATAATTTTATATATTTTATTTTAAATTTTATTTTAA 128614

Qy 230 agaatttctg-attctttaaataatgagatataatccttcttgatgtaaggtca 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128615 AATATATATAATTTTATTTATTAATAATTAATTTTAAATTTTATTAATATATATTT 128674

Qy 289 attcttctgctatgctttaaataaacaatgagccttactctctgtaagataaaaa 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128675 ATATATATATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 128734

Qy 349 agattatctcttcttgaagggcgtatcttaaagagcaagtgataaatcaattt 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128735 ATTAATATTTTATTTTAAATTTTAAATTAATTAATTAATTAATTTTATTTT 128794

Qy 409 acttaataaataaataaacaagaaatttataaagaattatataaaggattt 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128795 AATTAATAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 128854

Qy 469 ttacatgaaaggaagcctaaagtaatt 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128855 TTATTAATATATATTAATTAATTAATTAATTT 128883

RESULT 13
CONSOLRGO    175229 bp      DNA      HTG      29-JUN-2001
LOCUS        Homo sapiens chromosome 14 clone R-828K24, *** SEQUENCING IN
DEFINITION   PROGRESS ***, 2 ordered pieces.
ACCESSION    AL159141
KEYWORDS     AL159141.4 GI:14588844
VERSION      HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT.
SOURCE       human
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 175229)
AUTHORS      Helling, R., Pettit, J.L., Vico, V., Dasilva, C., Robert, C., Mincker, P.,
              Brottier, P., Catolico, L., Barbe, Y., Pelletier, E., Artiguenave, F.,
              Levy, M., Eckenberg, R., Bruls, T., Deberardinis, V., Cruaud, C.,
              Gyapay, G., Saurin, W. and Weissensbach, J.
              Sequencing of the human chromosome 14
              Unpublished
              2 (bases 1 to 175229)
              Genoscope.
              Direct Submission
              Submitted (29-JUN-2001) Genoscope - Centre National de Sequencage :
              BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)

```

```

COMMENT
On Jul 3, 2001 this sequence version replaced g1:10944754.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: Sequef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast vector,
phage, etc. . . even if efforts are made to eliminate these
contaminating sequences.
Upstream BAC (overlapping the T7 end) : R-618B21
Downstream BAC (overlapping the SP6 end) : R-807G16 (AC-AL136332)
----- Summary Statistics -----
Assembly program: Phrap; version 2.0
Quality coverage: 6.28x in 920 bases; sum-of-contigs
-----
Contigs composition :
71802 bp contig from 1 to 71802
103327 bp contig from 71903 to 175229
-----
Overall quality chart :
Range      : bases
0           : 363
1 - 9      : 12
10 - 19    : 99
20 - 29    : 216
30 - 39    : 714
40 - 49    : 2946
50 - 59    : 6725
60 - 69    : 13088
70 - 79    : 33326
80 - 89    : 64652
90 - 99    : 53078
-----
Percentage of bases with a quality value >= 40 : 99 %.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 71802: contig of 71802 bp in length
* 71803 71902: gap of 100 bp
* 71903 175229: contig of 103327 bp in length.
Location/Qualifiers
1..175229
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone_1lb="R-828K24"
 /clone_1lb="RP11-11"
 /note="matching EMBL:M79260
Rdbb:RH28481
Rdbb:RH28423
dbSTS:STS9346
Identified using the e-PCR software (G. Schuler)"
110596..110728
 /note="matching EMBL:M52863
Rdbb:RH45925
dbSTS:STS38993
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 44138 a 41136 c 42653 g 47199 t 103 others
ORIGIN

```













GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 23:30:59 ; Search time 311.54 Seconds  
(without alignments)  
40.132 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719  
Sequence: 1 MKRKAKILFLLSTVLPAQ.....FQNFVEKIESEKPESSPKN 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1701	99.0	341	1	FLAA_BORBU
2	403	23.4	337	1	FLAA_BORBU
3	361.5	21.0	350	1	FLAA_SPIAU
4	331.5	19.3	320	1	FLAA_TREHY
5	109	8.3	444	1	YCXD_BACSU
6	108	6.3	580	1	Y686_METUA
7	106.5	6.2	695	1	VARI_METUA
8	105.5	6.1	705	1	CMBA_BACSU
9	105	6.1	902	1	ITHL_PTC
10	104	6.1	499	1	XIIB_BACSU
11	103	6.0	650	1	ITHL_LYCES
12	102.5	6.0	914	1	ITHL_MESAU
13	101.5	5.9	613	1	CGAA_CLOBI
14	101.5	5.9	622	1	YF54_METUA
15	101.5	5.9	676	1	HS7C_TRYRB
16	101	5.9	1682	1	MSPL_PLAFJ
17	100	5.8	701	1	SYGB_HELPJ
18	99.5	5.8	907	1	ITHL_MOUSE
19	98.5	5.7	416	1	P46_MYCH
20	98.5	5.7	644	1	HS72_LYCES
21	98	5.7	707	1	YJ9C_YEAST
22	97.5	5.7	721	1	ENP1_TORCA
23	96	5.6	1875	1	MLP1_YEAST
24	95.5	5.6	434	1	ENO_IDLPE
25	95.5	5.6	631	1	HS71_ARATH
26	95.5	5.6	931	1	DAP1_YEAST
27	95.5	5.6	1104	1	BUD2_YEAST
28	95	5.5	641	1	HS74_YEAST
29	95	5.5	1024	1	STYP_STRAU
30	95	5.5	1125	1	MPD_BORBU
31	95	5.5	1701	1	MSPL_PLAFJ
32	95	5.5	1701	1	MSPL_PLAFM
33	94.5	5.5	680	1	OPDA_ECOLI

34	94	5.5	655	1	HS70_DAUCA	P26791	daucus caro
35	93.5	5.4	406	1	RENI_HUMAN	P00797	homo sapien
36	93.5	5.4	494	1	AMY1_SACFI	P21567	saccharomyc
37	93.5	5.4	651	1	HS7C_PETHY	P09189	petunia hyb
38	93.5	5.4	1672	1	PMPP_CHLAM	O9PJ32	chlamydia m
39	93	5.4	282	1	RUIA_HUMAN	P09012	homo sapien
40	93	5.4	282	1	RUIA_XENLA	P45429	xenopus lae
41	93	5.4	341	1	MDHM_BRANA	O43744	brassica na
42	93	5.4	454	1	MURC_AQUAE	O67373	aquifex neo
43	93	5.4	608	1	KU70_HUMAN	P12956	homo sapien
44	93	5.4	701	1	SYGB_HELPJ	P56454	helicobacte
45	93	5.4	1296	1	ASAI_ENTFA	P17953	enterococcu

## ALIGNMENTS

RESULT 1  
ID FLAA\_BORBU STANDARD; PRT; 341 AA.  
AC P70856; O51612; Q44876;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).  
GN FLAA OR BB0668.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_Taxid=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=212;  
RX MEDLINE=97144545; PubMed=8990312;  
RA Ge Y., Charon N.;  
RT "An unexpected flaa homolog is present and expressed in Borrelia burgdorferi."  
RT J. Bacteriol. 179:552-556(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35210 / B31;  
RX MEDLINE=98065943; PubMed=9403685;  
RA Fraser C.M., Casjens S., Huang W.M., Hickey E.K., Gwin M., Lathigra R., White O., Ketchum K.A., Dodson R., Richardson D., Dougherty J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J., Utechtack T., Matthey L., McDonald L., Arlath P., Bowman C., Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B., Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."  
RT Nature 390:580-586(1997).  
RN [3]  
RP SEQUENCE OF 276-341 FROM N.A.  
RC STRAIN=212;  
RX MEDLINE=9848936; PubMed=9765799;  
RA Old I.G., Trueba G.A., Saint-Girons I., Johnson R.C.;  
RT "A chea chew operon in Borrelia burgdorferi, the agent of Lyme disease."  
RL Res. Microbiol. 148:191-200(1997).  
CC -!- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.  
CC -!- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS SEVERAL ANTIGENICALLY RELATED POLYPEPTIDES.  
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC -----
DR EMBL; U62900; AAC44770.1;
DR EMBL; AE001168; AAC67025.1; ALT_INIT.
DR EMBL; X91907; CA63001.1;
DR TIGR; BB0668;
KW flagella; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 341
FT CONFLICT 254 255 EQ -> VK (IN REF. 1).
FT CONFLICT 317 318 MP -> IA (IN REF. 1 AND 3).
SQ SEQUENCE 341 AA; 38441 MW; 60B1475B8ADF3451 CRC64;

Query Match 99.0%; Score 1701; DB 1; Length 341;
Best Local Similarity 98.8%; Pred. No. 3.5e-117;
Matches 337; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRKAKSLIFLLSTVLAQETDGLAEGSKRAPEGLVDFAELARDPSRFLDTNLYVD 60
DB 1 MKRKAKSLIFLLSTVLAQETDGLAEGSKRAPEGLVDFAELARDPSRFLDTNLYVD 60
QY 61 VYSGASGIVPEDMVVDLGINNMSVLLTPSARLQAYKNSVAVPAVYKSESKRYAGDTI 120
DB 61 VYSGASGIVPEDMVVDLGINNMSVLLTPSARLQAYKNSVAVPAVYKSESKRYAGDTI 120
QY 121 LGVRVLPFSYSSQSSAMIMPPFKIPYSGESGNOFLGKLIDIKTKMKIKVSVSLGVEI 180
DB 121 LGVRVLPFSYSSQSSAMIMPPFKIPYSGESGNOFLGKLIDIKTKMKIKVSVSLGVEI 180
QY 181 DLEVLFEDMNGMEYAVSMGTLEFKGMADLIMSNPNYIPNISRRIKDDVPNPPLASSKMR 240
DB 181 DLEVLFEDMNGMEYAVSMGTLEFKGMADLIMSNPNYIPNISRRIKDDVPNPPLASSKMR 240
QY 241 FKAFVSKSHSKVKNFFFYKDLRVLDKLSVSDSDSESVKRYETSGTESLRKLK 300
DB 241 FKAFVSKSHSKVKNFFFYKDLRVLDKLSVSDSDSESVKRYETSGTESLRKLK 300
QY 301 AHEFFKRVKLREKISIAEGSFONFEKIESEKPESSPKN 341
DB 301 AHEFFKRVKLREKISIAEGSFONFEKIESEKPESSPKN 341
DB 301 AHEFFKRVKLREKISIAEGSFONFEKIESEKPESSPKN 341

RESULT 2
FLAA_SPIAU STANDARD; PRT; 337 AA.
AC P21982;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).
GN FLAA.
OS Spirochaeta aurantia.
OC Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
OX NCBI_TaxID=147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89155480; PubMed=2921247;
RA Brimhamsha B., Greenberg E.P.;
RT "Cloning and sequence analysis of flaa, a gene encoding a Spirochaeta
  aurantia flagellar filament surface antigen.";
RL J. Bacteriol. 171:1692-1697(1989).
RN [2]
RP REVISIONS.
RA Greenberg E.P.;
RL Submitted (JAN-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 22-49.
RC STRAIN=M1;
RX MEDLINE=91123217; PubMed=1991729;
RA Parales J. Jr., Greenberg E.P.;
RT "N-terminal amino acid sequences and amino acid compositions of the
  Spirochaeta aurantia flagellar filament polypeptides.";
RL J. Bacteriol. 173:1357-1359(1991).
```

```
CC -----
CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC -1- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF
CC REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS ONE OR ALL OF
CC FIVE ANTIGENICALLY RELATED POLYPEPTIDES.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24459; AAA9273.1;
DR PIR; A32814; A32814.
KW flagella; Periplasmic; Signal.
FT SIGNAL 1 21
FT CHAIN 22 337
SQ SEQUENCE 337 AA; 36857 MW; 3326E4B096785635 CRC64;

Query Match 23.4%; Score 403; DB 1; Length 337;
Best Local Similarity 30.4%; Pred. No. 2e-22;
Matches 100; Conservative 67; Mismatches 124; Indels 38; Gaps 9;

QY 1 MKRKAKSLIFLLSTVLAQETDGLAEGSKRAPEGLVDFAELARDPSRFLDTNLYVD 60
DB 1 MKRKAKSLIFLLSTVLAQETDGLAEGSKRAPEGLVDFAELARDPSRFLDTNLYVD 60
QY 61 VYSGASGIVPEDMVVDLGINNMSVLLTPSARLQAYKNSVAVPAVYKSESKRYAG 117
DB 61 VYSGASGIVPEDMVVDLGINNMSVLLTPSARLQAYKNSVAVPAVYKSESKRYAG 117
QY 118 DLEVLFEDMNGMEYAVSMGTLEFKGMADLIMSNPNYIPNISRRIKDDVPNPPLASSKMR 167
DB 118 DLEVLFEDMNGMEYAVSMGTLEFKGMADLIMSNPNYIPNISRRIKDDVPNPPLASSKMR 167
QY 167 STQINILGRNYINRLSLLEQNGDEREIVMGYLFDMGKSLQWNNPNYQTEVRRRDQ 225
DB 167 STQINILGRNYINRLSLLEQNGDEREIVMGYLFDMGKSLQWNNPNYQTEVRRRDQ 225
QY 228 DVPNPPLASSKMRFAFRVSKSHSKVKNFFFYKDLRVLDKLSVSDSDSESVKRYETSGTESLRKLK 284
DB 228 DVPNPPLASSKMRFAFRVSKSHSKVKNFFFYKDLRVLDKLSVSDSDSESVKRYETSGTESLRKLK 284
QY 285 -----FKVYETSGTESLRKLKAHETFK 306
DB 285 -----FKVYETSGTESLRKLKAHETFK 306
DB 286 LRQREQYRNFELAKLGNLQVRSLEKRR 314

RESULT 3
FLAA_TREPA STANDARD; PRT; 350 AA.
AC P18193;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).
GN FLAA OR TP0249.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90307197; PubMed=2194955;
RA Isaacs R.D., Radolf J.D.;
RT "Expression in Escherichia coli of the 37-kilodalton endoflagellar
  sheath protein of Treponema pallidum by use of the polymerase chain
  reaction and a T7 expression system.";
RL Infect. Immun. 58:2025-2034(1990).
RN [2]
RP SEQUENCE FROM N.A.
```

```

RC STAIN-NICHOLS: PubMed-9665876;
RX MEDLINE-98332770: pubmed-9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dooson R., Gilm M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khakhal H., Richardson D., Howell J.K., Chlamdaram M., Uteckack T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RT Science 281:375-388(1998).
RL 131
RP SEQUENCE OF 30-350 FROM N.A.
RC STAIN-NICHOLS:
RX MEDLINE-90035409: pubmed-2680972;
RA Isaacs R.D., Hanks J.H., Guzman-Verduzco L.-M., Newport G.,
RA Agabian N., Norgard M.V., Lukehart S.A., Radolf J.D.;
RT "Molecular cloning and DNA sequence analysis of the 37-kilodalton
RT endoflagellar sheath protein gene of Treponema pallidum."
RL Infect. Immun. 57:3403-3411(1989)
CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC -1- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF
CC REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS SEVERAL
CC ANTIGENICALLY RELATED POLYPEPTIDES.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63142; AAA27477.1;
DR EMBL: AE001206; AAC65235.1;
DR EMBL: M26525; AAA27476.1;
DR TIGR: TP0249;
KW Flagella; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 350 FLAGELLAR FILAMENT OUTER LAYER PROTEIN.
SQ SEQUENCE 350 AA; 38857 MW; 2BC7CFEA0CDBA4C7 CRC64;

Query Match 21.0%; Score 361.5; DB 1; Length 350;
Match Local Similarity 28.5%; Pred. No.2.e19;
Matches 100; Conservative 66; Mismatches 120; Indels 65; Gaps 11.

QY 6 KSIL--FELLSTVLPFAQGTDLGAESGSKRAEPDELVDFAELARDPSSRFLDLTNYVDVY 63
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 22 ESVLIDFAKLNAVDIADKSGGTHRR-----TYLDIASLAD-----TSTTD--- 63
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 64 SGASGIVKPEDMVVDLGINNVSLLTPSARLQAYVKNVAVAPVYKSESRYAGDTTIGV 123
Db -----EOKALMRSSSLAAOWMEVVLINSARNPVAAHSAVRIBAPVSEKASPAGERVIGV 117
QY 124 RVLPFSY--SOSAMIMPFKIPFY-----GSGN-----QL-GLGLDIN 162
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 RVLPFTWDSNNAMKPPAFVLPVAYVMAQVDDQVNAVAPTEEEKASGKGRFEDGQVKN 177
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 163 IKTMKEIVSVYSLQLEIDLLEVLFDPMGMEVAYSMGLTKFKGMDLINSFNYPINIS 222
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 178 VGVLSIAVNNYVGNMYPGLGVYMMMDQGEVIRYFMGYLLFPMSKELVANNNSYISDVNS 237
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 223 RIKDDVNPYPLASSKMFKAFRVSKSHSSKVKNFIFYVKDLRVLYDKLVSIDSIDISE 282
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 238 REVR-LYVUYVPASTRPVHVVVEGFMVTRDAHAAGDVGVGFQDKIITYDKRAVLSTVDEADE 296
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 283 SVFKVYTERSGSLRKLKAHEFFKVLVLRKISIAESGFQVFKERISEK 333
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 DLM-----GIQARRAE-----RKVEVARRGQGVQLVRIEDQK 330
Db :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 4
ID FLAI_TREHH STANDARD: PRT: 320 AA.
AC P32520; P80157;
AD 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN FLAI1 PRECURSOR (44 KDA SHEATH
DE PROTEIN).
GN FLAI1 OR FLAN.
OS Treponema hyodysenteriae (Serpulina hyodysenteriae).
OC Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
OX NCBI_TaxId=159;
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-38.
RC STRAIN=C5;
RX MEDLINE=92307926; PubMed=1612759;
RT "Koopman M.B.H., de Leeuw O.S., van der Zeijst B.A.M., Kusters J.G.;"
RT "Cloning and DNA sequence analysis of a Serpulina (Treponema)
RT hyodysenteriae gene encoding a periplasmic flagellar sheath
RT protein."
RL Infect. Immun. 60:2920-2925(1992).
RN [2]
RP SEQUENCE OF 20-38.
RC STRAIN=C5;
RX MEDLINE=93139764; PubMed=1487733;
RA Koopman M.B.H., Baats E., van Vorstenbosch C.J.A.H.V.,
RA van der Zeijst B.A.M., Kusters J.G.;"
RA "The periplasmic flagella of Serpulina (Treponema) hyodysenteriae are
RT composed of two sheath proteins and three core proteins."
RT J. Gen. Microbiol. 138:2697-2706(1992).
CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC -1- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF TWO
CC SHEATH PROTEINS, FLAI1 (44 KDA) AND FLAI2 (35 KDA) AROUND A CORE
CC THAT CONTAINS THREE PROTEINS FLAI3 (37 KDA), FLAI2 (34 KDA) AND
CC FLAI3 (32 KDA).
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; X63006; CAA44735.1; -.
DR EMBL; A24079; CAA01715.1; -.
DR PIR; A43824; A43824.
KW Flagella; Signal; Periplasmic.
FT SIGNAL 1 19
FT CHAIN 20 320
FT FT FLAGELLAR FILAMENT OUTER LAYER PROTEIN
FT FLAI1.
FT CONFLICT 21 21 T->S (IN REF. 2).
SQ SEQUENCE 320 AA; 35986 MW; 3CFC27B07F8646B4 CRC64;

Query Match 19.3%; Score 331.5; DB 1; Length 320;
Best Local Similarity 28.8%; Pred. No. 3e-17;
Matches 90; Conservative 63; Mismatches 129; Indels 31; Gaps 8;

OY 9 LEFLLSTVLAEQETGGLGEGSKRAPEGLVDLDFEAELARDPSSRLDTFTYVDVYVSGASG 68
DB 4 LEVVLTSIFIAASAVGLNST-----LIDFA-----LTGMADNLQAGEED 43

OY 69 IVKPEDMVVDLGINNMSVLLTPSARLQAVYKNSVAAPAVKSESKRYAGDTLLGVAVLEP 128
DB 44 TNEVPAVLENTLYNDWMVWMLNESARLTENRRRSYTVTNDKSCNMGAMEAGKVLGVRHFP 103

OY 129 SYSQSS-AMIMPRRIIPYSGSGNQL-GRGLDINIKTMKIKSVYSLGEIDLEVLV 186
DB 104 LAAMSVYALVYKRYVLEEMGADGKRYTEGKVIHNVGKISISSWYGRNLLISFYVNL 163

```

QY 187 EDMNGEAYSMGTLKFKGADLMSNPYIPNISRIKIDVNPYPLASSKMFKAERV 246  
 DB 164 QNEGELKSTPMGIVYNGRQVMEKREILPNRDLVIRE-PLYRMPISVLDISGF 222  
 QY 247 SKSHSSKVNKEIFYKDLRVLYDKLSVSDSIDSESVFKYETSGTESLKLKAHETFK 306  
 DB 223 YRTKDTKGFITVYKDVTELYDVVYVDFEEDIDDEATWOLK--TENDRK-QAIES-- 276  
 QY 307 RVLKREKISIAE 319  
 DB 277 --AIRFOAELRD 287  
 RESULT 5  
 YCXL\_BACSU STANDARD; PRT; 444 AA.  
 ID YCXL\_BACSU  
 AC 008792;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 CN HYPOTHETICAL 50.8 KDA PROTEIN IN SRFA4-SFP INTERGENIC REGION (ORF8).  
 YCXL  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RC MEDLINE=93360813; PubMed=8355609;  
 RA Cosmina P., Rodriguez F., de Ferris F., Grandi G., Perego M.,  
 RA Venema G., van Sinderen D.;  
 RA "Sequence and analysis of the genetic locus responsible for surfactin  
 RT synthesis in Bacillus subtilis";  
 RT Mol. Microbiol. 8:821-831(1993).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=168;  
 RX MEDLINE=97124189; PubMed=8969502;  
 RA Yamane K., Kumano M., Kurita K.;  
 RA "The 25 degrees-36 degrees region of the Bacillus subtilis chromosome:  
 RT determination of the sequence of a 146 kb segment and identification  
 RT of 113 genes";  
 RL Microbiology 142:3047-3056(1996).  
 CC CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (POTENTIAL).  
 CC CC -1- SIMILARITY: IN THE C-TERMINAL, TO AMINOTRANSFERASES.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X70356; CAA49821.1; -  
 DR EMBL: D50453; BAA08990.1; -  
 DR EMBL: 299105; CAB12150.1; -  
 DR EMBL: 299106; CAB12164.1; -  
 DR Subtilisin; BG10175; YCXL.  
 DR InterPro: IPR000524; HTH\_GNTR.  
 DR Pfam: PF00392; gntr; 1.  
 DR PRINTS: PR00035; HTHGNTR.  
 DR SMART: SM00345; HTH\_GNTR; 1.  
 KW Hypothetical protein; Transferase; Aminotransferase;  
 FT Pyridoxal phosphate; Complete proteome.  
 FT BINDING 286 Complete PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 444 AA; 50755 MW; 1C392C3D386696E7 CRC64;

Matches 75; Conservative 42; Mismatches 126; Indels 104; Gaps 15;  
 QY 29 SKRAPEGLVDFAEFLARDPSSRDLNRYD-----YYSGASGI--- 69  
 DB 73 SKSGQPGP--IDFATSDAPD-----DVFPIYDFQHCINKALIDYTKNDLFITGPKGLPSL 125  
 QY 70 -----VKPEDMVVDLGINNSVYL-----TPSARLQAVYKN 100  
 DB 126 IRLVRLKLLATQGVAFDRRHIFITSGVOALSLCAMPFPNGKEKIALIEQGYHLMBQLE 185  
 QY 101 SYVAPAVYKSSKRYACDTLIGVAVLPFSYSQSSAMIMPPKIPFYGSEGNQPLGGLI 160  
 DB 186 TLGIPAVGVKRTKE--GLDIKVERLFPOTESIKFFYTPRPHN-----LGCSLS 233  
 QY 161 DNITKMEIKVSYSLGYEIDLEVLFEEDMNGMEYASMGTLKFKGADLMSNPYIPNI 220  
 DB 234 ERDKQ-----ELVRLAEANDVYLYVEDDY-----LGDLLENKKAD-----PLYAYDL 274  
 QY 221 SSRIRKIDVNPYPLASSKMFKAERVSKSHSKVNFYVYKDLRVLYDKLSVSDSID 280  
 DB 275 SSHYI-----YLSFKSMFPGRLVGAVALPEALDTDTFYA-----YKKLN-DIDCSMI 321  
 QY 281 SESVFKYETSGTSLKLKAHETFK-RVLKREKI-----SIAGSF 322  
 DB 322 SQALLETLYKSGMYGRHREKIRDSYKERSLRLHQAIRTHROLGSGRF 368

RESULT 6  
 Y686\_METJA STANDARD; PRT; 580 AA.  
 ID Y686\_METJA  
 AC Q58099;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 CN HYPOTHETICAL PROTEIN M0686.  
 DE M0686.  
 GN Methanococcus jannaschii.  
 OS Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrman J.L., Nguyen A.,  
 RA Udenback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii";  
 RL Science 273:1058-1073(1996).  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U67515; AAB98681.1; -  
 DR TIGR: M0686; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 580 AA; 68392 MW; C23CA146575B2506 CRC64;

Query Match 6.3%; Score 109; DB 1; Length 444;  
 Best Local Similarity 21.6%; Pred. No. 0.83;

Query Match 6.3%; Score 108; DB 1; Length 580;  
 Best Local Similarity 21.7%; Pred. No. 1.4;  
 Matches 56; Conservative 38; Mismatches 66; Indels 98; Gaps 11;

```

QY 112 SKRYAGDTILGVRLFPSSYSSAMIMPKRIPYSGESGNQFLGKLDINIKTKREIKV 171
DB 376 SERISGST-----ESKMKDKREVSQWIENNKKI--PEVEVLEKI--EIQY 417
QY 172 SVSLGCEIDLEVLFEEDMNGMEYAYSMGTLKFGKMGADLIWSPN---YIPNISRIIKDD 228
DB 418 GAVSL-----KKYKSSGSSKRYGVNLIKRNKPMDYKP-----DN 453
QY 229 VPNYPLASSMRFKAF-----RVSKSHSSKVKNFIEYVK 262
DB 454 IAYKLEDDHHIFPKGFLRNKGISNEYIDYLNKTPILDETNNKISKKSPSK-----YVK 507
QY 263 DLAVLKDLSVSDSD-----SESVFKYVETSGTESLRKLKAHEFRKRLVLT 311
DB 508 EMEIQKNNKGLSDEAVNKKVEILKGHINDEMEFLRNT--DLSLKDIEENFNRIEL 566
QY 312 REKISIEGSGFQNFVEKI 329
DB 567 REKL-----ILEKI 575

RESULT 7
VATILMETUA STANDARD: PRT: 695 AA.
AC 057675;
DF 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
V-TYPE ATP SYNTHASE SUBUNIT I (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT I).
GN ATP1 OR MJ0222.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1/DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sulton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Peterson J.F., Fuhrman J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Weidman J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67478; AAB98208.1; -.
CC TIGR: M02222; -.
DR InterPro: IPR002490; V_ATPase_sub_a.
DR Pfam: PF01496; V_ATPase_sub.a; 1.
KM Hydrolase: Hydrogen ion transport; Transmembrane; Complete proteome.
FT TRANSMEM 392 412 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
FT TRANSMEM 497 517 POTENTIAL.
FT TRANSMEM 534 554 POTENTIAL.
FT TRANSMEM 556 576 POTENTIAL.
FT TRANSMEM 605 625 POTENTIAL.

```

```

FT TRANSMEM 627 647 POTENTIAL.
SQ SEQUENCE 695 AA; 76953 MW; F03E5CBE29D53D CRC64;

Query Match
Best Local Similarity 23.5%; Score 106.5; DB 1; Length 695;
Matches 84; Conservative 41; Mismatches 108; Indels 125; Gaps 19;

QY 68 GIVPEDMVVDLGINNSVLLTPSARLQAYKNSVAPAVYKSSKRYAGDTILGVRLV 126
DB 32 GIVELCDLSEKLEDELEKTLSPSSAD-YVRN--VTSIMLK-----AG-----RLLD 76
QY 127 -FSPYSQSSAMI-----MPKRIPEYSGESGNQFLGKLDINIKTKREIKV 161
DB 77 MESSVSQKETSIDKILNPKPEKKVSPNSYQEVYDAEAKVLNEISKEVDPAPRLSELD 136
QY 162 NIKT-MKEIKVSV-YSLGCEIDLEVLFEEDMNGMEYAY-----SMGTLKFGKMGADLI 210
DB 137 NKKSKLLQKQIYLYKGLFEDLKYL-----GSGEYVFAGASVPKREKGLK-----AELD 188
QY 211 WSNPNYIPNIS-SRIKDVDPNYPPLASSKMFKAFRVSKSHSSKVKNFIEYVKDLRVLYD 269
DB 189 KVADGYIGIFSGSEFEKDKIRVPIVFTLTKEKLENV-----LSEIRKFEFERDYI----- 239
QY 270 KLSVSDSD-----ESVFKYVETSGTESLRKL--AHETFRVLKREKISI- 317
DB 240 -----SDVEGTPSEALSKIESLKAIESRNSLIEKTLAKLAKWEKELLAVYELLSTE 292
QY 318 -----AEKSFQNF-----EVEKIESEKPESSP 339
DB 293 KARGDAVSGFCKTRTYIEMVAPARDAEKAKSLIENSADGFARVEITPEPEPEKIP 350

RESULT 8
CMBB_BACSU STANDARD: PRT: 705 AA.
AC 002113;
DF 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
V-TYPE ENHANCER PRECURSOR (MODIFIER PROTEIN OF MAJOR AUTOLYSIN).
GN LYTB OR CMBB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93018998; PubMed=1357079;
RA Lazarevic V., Margot P., Soldo B., Karamata D.;
RT "Sequencing and analysis of the Bacillus subtilis lytRBC divergon: a
RT regulatory unit encompassing the structural genes of the N-
RT acetylmutamoyl-L-alanine amidase and its modifier."
RL J. Gen. Microbiol. 138:1949-1961(1992).
CC [2]
CC SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
CC Kuroda A., Rashid H.M., Sekiguchi J.;
CC MEDLINE=92407479; PubMed=1356138;
CC "Molecular cloning and sequencing of the upstream region of the major
CC Bacillus subtilis autolysin gene: a modifier protein exhibiting
CC sequence homology to the major autolysin and the spoliid product."
CC J. Gen. Microbiol. 138:1067-1076(1992).
CC -1- FUNCTION: POSSIBLY INVOLVED IN CELL WALL METABOLISM DURING SPORE
CC FORMATION. ENHANCES THE AMIDASE ACTIVITY APPROXIMATELY THREEFOLD.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- SIMILARITY: WITH THE N-TERMINAL OF LYTC AT ITS N-TERMINAL AND THE
CC B. SUBTILIS SPOIID GENE AT ITS C-TERMINAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M87645; AAA22580.1; -  
DR EMBL: D10388; BAA01224.1; -  
DR EMBL: Z99122; CAB15580.1; -  
DR PIR: C47679; C47679.  
DR Subtilist: BG10406; lytB.  
KW Membrane; Signal; Cell wall; Repeat; Complete proteome.  
FT SIGNAL 1 25  
FT CHAIN 26 705 AMIDASE ENHANCER.  
FT DOMAIN 64 349 3 X TANDEM REPEATS.  
FT REPEAT 64 161 1.  
FT REPEAT 162 251 2.  
FT REPEAT 252 349 3.  
SQ SEQUENCE 705 AA; 76729 MW; A1A2449C2D801528 CRC64;

Query Match 6.1%; Score 105.5; DB 1; Length 705;  
Best Local Similarity 25.1%; Pred. No. 2.8; Mismatches 120; Indels 75; Gaps 14;  
Matches 76; Conservative 32; Mismatches 120; Indels 75; Gaps 14;

QY 46 RPPSSTRDLT-NYYDYVSGASGIVKPEDMVVDGINMWSVLLPSPARLQAYKNSVY- 103  
DB 252 RPPGSTRYELTANIIKOLK-----LKADKVVMTNCTKYADVLIGASL---ASKKNSQL 302  
QY 104 -----APAVKSESK---RYAGDTILSYRVLPSPSOSAMIMPFKIPFSGEGNPF 154  
DB 303 EYKQDSVPPAAKSTIRKATYAVDFI-----GSTSSISAEEVNSLADEFLADGTVN 355  
QY 155 L-----GKGLIDNIKT---MKEIKVSYSLGEIDLEVLFEEDMNGMEYVSMGLKFGWA 207  
DB 356 LKINSKMLKLENIKTGNSLRKIPENYTSNRIISL-----GKQY---LGVVNS--- 402  
QY 208 DLWSPNPIYIPNISRI-----IKDDVPN-YPLASSKMKFKAFFRSKSHSKVNFIFYV 261  
DB 403 ---IESTKYIRPVNENIPEDYKLVIPNEMPASMSLEAKQTVARVYITKT----- 454  
QY 262 KDLRVLYDKLVSISDDISSEVFKYETSGTESLKLKAHEFKVVLKREISIAEGS 321  
DB 455 -----GTVVDTTAFQVYGYGYSWNSNTNKAVEQTKGKVLKNSGLITAAYS 500  
QY 322 FON 324  
DB 501 SSN 503

RESULT 9  
ITHL\_PIG STANDARD: PRT; 902 AA.  
AC 029052; 002669;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-JUG-2001 (Rel. 40, Last annotation update)  
DE INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN H1 PRECURSOR (ITI HEAVY  
DE CHAIN H1).  
GN ITIHL.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE=Liver;  
RA Suzuki H., Hamasima N., Kimura M., Ozawa A., Yasue H.;  
RN Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.  
RP [2]  
RP SEQUENCE OF 124-902 FROM N.A.  
RA TISSUE=Liver;  
RA Gebhard W.;  
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.  
CC -!- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A

BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,  
INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE  
LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE  
ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY  
SIMILARITY).  
-!- SUBUNIT: I-ALPHA-1 PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM  
ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN,  
BIKUNIN. INTER-ALPHA-INHIBITOR (I-ALPHA-1) IS COMPOSED OF H1, H2  
AND BIKUNIN. INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND  
BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-1) OF H3 AND BIKUNIN (BY  
SIMILARITY).  
-!- P1M: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN  
4-SULFATE BRIDGE TO THE THEIR C-TERMINAL ASPARATE (BY  
SIMILARITY).  
-!- SIMILARITY: BELONGS TO THE ITI FAMILY.  
-!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D38754; BAA07632.1; -  
DR EMBL: Y11546; CAA72309.1; -  
DR InterPro: IPR002035; VWFA.  
DR Pfam: PF00092; vwa; 1.  
DR SMART: SM00327; vwa; 1.  
DR PROSITE: PS50234; VWFA; 1.  
KW Serine protease inhibitor; Repeat; Signal; Multigene family;  
KW Glycoprotein.  
FT SIGNAL 1 28  
FT CHAIN 29 663  
FT FT  
FT PROPEP 664 902  
FT DOMAIN 282 442  
FT CARBOHYD 69 66  
FT CARBOHYD 277 277  
FT CARBOHYD 741 741  
FT BINDING 663 663  
FT FT  
FT CONFLICT 432 432  
FT CONFLICT 439 439  
FT CONFLICT 456 456  
FT CONFLICT 574 574  
FT CONFLICT 712 712  
FT CONFLICT 777 777  
FT CONFLICT 779 780  
FT CONFLICT 793 793  
FT CONFLICT 823 823  
FT CONFLICT 836 836  
FT CONFLICT 893 893  
SQ SEQUENCE 902 AA; 100395 MW; BD950C230FF14185 CRC64;

Query Match 6.1%; Score 105; DB 1; Length 902;  
Best Local Similarity 19.7%; Pred. No. 4.2; Mismatches 118; Indels 106; Gaps 17;  
Matches 70; Conservative 61; Mismatches 118; Indels 106; Gaps 17;

QY 12 LLSVLFPAQETGLAEGSRARPEGLVLDFAELARDPSSTRDLNYYVYVSGASGIYK 71  
DB 281 LKNVVFVIDISSMEGQVAKFKELKIL-----SDLKPGDYEDLVLFSGA----- 328  
QY 72 PEDMVVDLGINW--SVLLTPSARLQA---YVKNVYVAVAVKSSKRYAGDTILGVRVL 126  
DB 329 -----VQSMRGSLVQASTANLDAARSYVROPSLA-----GSTNLNGSLGIELL 373  
QY 127 -----FPSYSSOSAMIMPFKIPYSGEGNPFGLKGLIDNIKTMKETKSVV-----Y 174  
DB 374 NKAQGLSFPFSNRASLIL-----IMLTGDP-----TEGVTDRSQILKVNRAIRGRFPLY 423





[illegible]

RT protein with homologues to *Bacillus thuringiensis* delta-endotoxins.";  
 RL J. Bacteriol. 178:3099-3105(1996).  
 CC -1- FUNCTION: TOXIN ACTIVE ON MOSQUITO LARVAE OF THE SPECIES Aedes  
 CC AEGYPTI, COLEX PIPRIENS, AND ANOPHELES STEPHENSI.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DEVELOPMENTAL STAGE: PRODUCED DURING SPOULATION. THE EXPRESSION  
 CC DECREASES CONSIDERABLY WITH CELL LYSIS.  
 CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X94146; CAA63860.1; -  
 DR Interpro: IPR001178; Endotoxin.  
 DR Pfam: PF00555; endotoxin; 1.  
 KM Toxin: Sporulation. 160  
 FT DOMAIN 155  
 FT CONFLICT 9  
 FT SEQUENCE 613 AA; 71172 MW; 410EFPD9C601CEE5 CRC64;  
 SQ  
 Query Match 5.9%; Score 101.5; DB 1; Length 613;  
 Best Local Similarity 18.8%; Pred. No. 4.5;  
 Matches 54; Conservative 59; Mismatches 97; Indels 77; Gaps 12;  
 QY 42 AELARDPSSRLDLTVNDVYSGASGIVKPEMDVLDGINNSVLTLPARLQAYKNS 101  
 DB 208 SNKLDPNNSWIDITRCRM-----TETIDM-----ISICPIYDKRYDK-- 249  
 QY 102 VPAVAVKSESKRYAGDTLIGVRLP-----SYSSSAMIMPPKIPYSGESGNQFL 155  
 DB 250 ---PIMNQTLTKRYSDPVNFIDENIPISYEKKYINISPELESTLFSISYTNKSGKFL 306  
 QY 156 G-----KGLIDNIKMKELKVSYSIGYFIDLEVLFEEDNGMEYAL-----SMG 199  
 DB 307 NGHVNRRHVGTDLVNCGIRETHYNGYSNVEE-SMAFDIDKAYSNNFNNTONNPTSVK 365  
 QY 200 TLKF---KGWADLIWSNPN-----YIPNISRIKIDVPVYPLASS-KMR--- 240  
 DB 366 SIFELTKKNDENIYGPDSNIDFTFRNIGYLSNLNNESTYSLSMILANNDKIQINT 425.  
 QY 241 -----FKAFVYSKSHSSKVKNFIFYKDLRYL--YDKLSV 273  
 DB 426 DTPHSYSYWIYKGIEDTNYISDKLIQIPLVKEVKLKSRYSEISV 472  
 RESULT 14  
 YF54\_METJA STANDARD; PRT; 622 AA.  
 AC 058949;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PROTEIN MJ1554 PRECURSOR.  
 GN MJ1554.  
 OS Methanococcus jannaschii.  
 OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 OX NCBI\_TaxId=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=868087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merriam J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,

RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: TO A Fulgidus AF0817.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U67596; AAB99574.1; -  
 DR TIGR: MJ1554; -  
 KM Hypothetical protein; Signal; Complete proteome.  
 FT SIGNAL 1  
 FT CHAIN 21 622  
 FT SEQUENCE 622 AA; 72282 MW; 86CA62F3854D73F CRC64;  
 SQ  
 Query Match 5.9%; Score 101.5; DB 1; Length 622;  
 Best Local Similarity 20.9%; Pred. No. 4.6;  
 Matches 77; Conservative 74; Mismatches 149; Indels 69; Gaps 20;  
 QY 1 MKRKAKSLPFLSTV-LFAQETDGLAEGSKRABGELVLDFAELARDPSSRLDLTVNV 59  
 DB 1 MKIAVAIFSLMLIISLFS---GCVENEKPIEGS--NDFKLIPVNSKNFEFKMTV 54  
 QY 60 D-----VYSGASGIVKPEMDVLDGINNSVLTLPARLQAYKNSVPAVAVKSESK 114  
 DB 55 ENSIGNIYVGHVSASREVOJTSTVKSNN---VETSEPERFSKTVNQVGVDEADILK 110  
 QY 115 YAGDTLIGVRLPFSYSSAMIMPPKIPYS-----GSGNQFLKGLIDNIKMKKE 168  
 DB 111 TNGNIT-----AFSQNKIYLIKPLP-PKYAIIKINISCGLYLTN--NFLIYISW 158  
 QY 169 IKVSVSLGYEIDLEVLFE-DMNGMEYAYSN--GTLKF--KGWAD-LIWSN----- 213  
 DB 159 NKITSYVNSNPENMKIIMQMDLNG-SYVDSRLYNGTLVYRKNSIDCPIVMNNYKIGYD 217  
 QY 214 PNTIPNISRIKIDVPVYPLASSKMKFKAFFVS-----KSHSKVKNFIFYKDL 264  
 DB 218 KYIPELPPYISMDPTTYII--SRINKSGKVENSIAYGVNKTYYMSNNLYFAYNL 275  
 QY 265 RVLDKLSVSDISDISESVKYEYETSGTESLRKKAHEFPKRVLKLREKISIAEGSPON 324  
 DB 276 KINEKKLML-----NFLNESADKYPTEYADKIKRVIENEDEGDNAKF---VEITE-TIER 327  
 QY 325 FVEKIESEK 333  
 DB 328 YLSLSPSEK 336  
 RESULT 15  
 HS7C\_TRYBB STANDARD; PRT; 676 AA.  
 AC P20030;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-MAY-1992 (Rel. 22, Last annotation update)  
 DE HEAT SHOCK COGNATE HSP70 PROTEIN.  
 OS Trypanosoma brucei brucei.  
 OC Eukaryota: Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxId=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90377290; PubMed=2398917;  
 RX Lee M.G.-S., Polvere R.I., van der Ploeg L.H.T.;  
 RA "Evidence for segmental gene conversion between a cognate hsp 70 gene

RT and the temperature-sensitively transcribed hsp70 genes of  
 RT Trypanosoma brucei.";  
 RL Mol. Biochem. Parasitol. 41:213-220(1990).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M32139; AAA30208.1; -  
 DR PIR: A45515; A45515.  
 DR HSP; P19120; INCT.  
 DR InterPro: IPR001023; HSP70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PROSITE: PS00297; HSP70\_1; 1.  
 DR PROSITE: PS00329; HSP70\_2; 1.  
 DR PROSITE: PS01036; HSP70\_3; 1.  
 DR ATP-binding; Heat shock; Multigene family.  
 KW ATP-binding; Heat shock; Multigene family.  
 SQ SEQUENCE 676 AA; 73694 MW; AD67B783124B785E CRC64;

Query Match 5.9%; Score 101.5; DB 1; Length 676;  
 Best Local Similarity 23.2%; Pred No 5.1; Indels 85; Gaps 16;  
 Matches 73; Conservative 44; Mismatches 112;

QY	63	YSGAGSIVKPEDMVVDLG-----INNMSVLLTPSARLQAYVKNVAP	105
DB	3	YEGALGI-----DLGTTSCVGVQWQNERVEIANDQGNRTTPS--YVAFVNEVLWG	52
QY	106	AVVKSSEKRYAGDTILGV-RVLEFPSQS--SAMTPPKTFPYGSEGN-----QFLG	156
DB	53	DAKSHAARGSGVLPFDARKRLIGRKFSDSVQSDMKHMPKVE--EGEKGAVMRYEHLG	110
QY	157	KGLIDIKTKMEIKVSVSLGEYIDLEVFEDMNGMEYAVSMCTLKFKGMADLIWSPNY	216
DB	111	EGML-----LQEQISARVLA--LKSCAFSLGKQVAKAVTV-----PAY	150
QY	217	IPNISRRIKD-----DVPNYPLASSKMKRKAFRVSKSHSKVKNFIIFYKDLRV	266
DB	151	FNDSGQKATKDGATAGLEVLRIINEPTAA-----IAYGLDKADEGKERNVL-----	198
QY	267	LYDKLSVSTSDSDSES--VKVYERSTESL--RKLKAHEFFKRYLKLREKISIAESGF	322
DB	199	VDFPGGCTPDVSIISVSGGVFEVKATNGDTHLGGEDVDALLLEHALADIRNRGIEGSL	258
QY	323	-QNFVEKIESEKPE	335
DB	259	SOKMLSKLRSCEE	272

Search completed: March 20, 2002, 00:06:00  
 Job time: 2101 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 01:58:06 ; Search time 2486.21 Seconds  
(Without alignments)  
139.345 Million cell updates/sec

Title: US-09-004-395-3

Sequence: 1 atgaaaggaagctaaagt 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb.ba: \*  
2: gb.htg: \*  
3: gb.in: \*  
4: gb.om: \*  
5: gb.ov: \*  
6: gb.pat: \*  
7: gb.ph: \*  
8: gb.pl: \*  
9: gb.pr: \*  
10: gb.ro: \*  
11: gb.sts: \*  
12: gb.sy: \*  
13: gb.un: \*  
14: gb.vi: \*  
15: em.ba: \*  
16: em.fun: \*  
17: em.hum: \*  
18: em.in: \*  
19: em.om: \*  
20: em.or: \*  
21: em.ov: \*  
22: em.pat: \*  
23: em.ph: \*  
24: em.pl: \*  
25: em.ro: \*  
26: em.sts: \*  
27: em.sy: \*  
28: em.un: \*  
29: em.vi: \*  
30: em.htg.hum: \*  
31: em.htgo.in: \*  
32: em.htgo.rod: \*  
33: em.htg.hum: \*  
34: em.htg.in: \*  
35: em.htg.rod: \*  
36: em.htg.other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	1655	1	BB062900
2	21	100.0	13271	1	AE001168
3	20	95.2	3885	8	LEPUR
4	19.4	92.4	183289	2	AC068117
5	19.4	92.4	198207	2	AC068539
6	18.4	87.6	33045	3	CET25C12
7	18.4	87.6	62555	2	AC091103
8	18.4	87.6	139064	2	AC073440
9	18.4	87.6	148142	9	AC011421
10	18.4	87.6	156048	2	AC091111
11	18.4	87.6	159012	2	AC023145
12	18.4	87.6	161624	2	AP001493
13	18.4	87.6	164652	2	AC026923
14	18.4	87.6	170270	2	AP001548
15	18.4	87.6	170413	2	AC090405
16	18.4	87.6	175178	2	AC011783
17	18.4	87.6	175850	2	AP001399
18	18.4	87.6	175924	2	AC092883
19	18.4	87.6	176421	2	AC025157
20	18.4	87.6	194487	2	AP002896
21	18	85.7	5954	10	AF247181
22	18	85.7	27697	2	AC017568
23	18	85.7	62238	2	AC012390
24	18	85.7	73360	9	HS06083
25	18	85.7	112621	9	AC004891
26	18	85.7	159570	3	AC010667
27	18	85.7	169572	2	AC023832
28	18	85.7	199891	9	CNS00M80
29	18	85.7	320565	5	AE003844
30	17.8	84.8	2794	5	XL019974
31	17.8	84.8	10255	2	AC015201
32	17.8	84.8	13164	1	AE006153
33	17.8	84.8	14733	2	AC014738
34	17.8	84.8	24137	9	AL157368
35	17.8	84.8	29034	8	HS243947
36	17.8	84.8	40397	8	SC9959
37	17.8	84.8	82289	8	ATAC009895
38	17.8	84.8	97912	9	HSJ753D10
39	17.8	84.8	111641	9	HS232N11
40	17.8	84.8	112755	10	AC083887
41	17.8	84.8	113253	2	AC092356
42	17.8	84.8	115884	2	AC017581
43	17.8	84.8	116518	2	AC067732
44	17.8	84.8	117453	3	AE003670
45	17.8	84.8	120314	9	AL133338

#### ALIGNMENTS

RESULT 1  
BB062900 1655 bp DNA BCT 15-JAN-1997  
LOCUS  
DEFINITION Borrelia burgdorferi flagellar filament outsheath protein (flaA)  
gene, complete cds, and chemotaxis histidine kinase (cheA) gene,  
partial cds.  
ACCESSION U62900  
VERSION U62900.1 GI:1575445  
SOURCE  
KEYWORDS  
ORGANISM Lyme disease spirochete.  
Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
REFERENCE 1 (bases 1 to 1655)  
Ge'Y. and Charon,N.W.  
TITLE An unexpected flaA homolog is present and expressed in Borrelia  
burgdorferi  
JOURNAL J. Bacteriol. 179 (2), 552-556 (1997)  
MEDLINE 97144545  
REFERENCE 2 (bases 1 to 1655)

AUTHORS Ge, Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1996) Yigong Ge, West Virginia University,  
Microbiology HSCN, Morgantown, WV 26506, USA  
FEATURES  
Source  
1. 1655  
/organism="Borrelia burgdorferi"  
/strain="212"  
/db\_xref="taxon:139"  
/clone="pm1 and pm2"  
1. 395  
/function="unknown"  
/note="oriA"  
/codon\_start=3  
/transl\_table=1  
/protein\_id="AAC4769.1"  
/db\_xref="GI:1575446"  
/translation="DNLEFSKRVYFPHSSKGVANLRYLRDQNLKDNLDLVKDFLLG  
SNEGSEFGLSDSRLFLSKNGVYVNLREFYDSFNNGDYNESFDVKVNLFPAMSL  
IKTRFNYPGRIKRIIIVEGILKEOS"  
473. 1498  
/gene="flaA"  
473. 1498  
/gene="flaA"  
/function="putative flagellar filament outsheath protein"  
/codon\_start=1  
/transl\_table=1  
/product="flaA protein"  
/protein\_id="AAC44770.1"  
/db\_xref="GI:1575447"  
/translation="WKRKASILFFLSTVLEPAETDGLAEGSKRAEPGEVLDFAE  
LRDPSSTRDLTNYVYVYSGASGVKPEDVVDGIMNWSYLLPSPARLQAYVNSV  
VAPAVKSESKRYAGDTLLGVLEFSPYSOSAMIMPEFKIPYEGESGNOFLGKI  
DNITMKRIKYSVSLGVEIDLEVLFEEDMGMEVYSGMTLFGKGMADLIMSNPYIP  
NISRSITKDDVPNPLASSKMRKAFRYSKSHSSKSKYKFIYVDLRYLVDKLSYD  
SDIDSESVFKYETSGETSLRKLKAHETFKYKLREKISTAESGFQNFVERKTESKP  
EESSEPKN"  
gene  
1566. 1655  
/gene="cheA"  
1566. >1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon\_start=1  
/transl\_table=1  
/product="CheA protein"  
/protein\_id="AAC44771.1"  
/db\_xref="GI:1575448"  
/translation="WEILDLENEELLGVFEPAQNLVDILEENT"  
BASE COUNT 560 a 182 c 308 g 603 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 1; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 atcgaaaggaagcctaaagt 21  
|||||  
DB 473 ATGAAAGGAAGCAAGCTAAAGT 493  
RESULT 2  
AE001168 13271 bp DNA BCT 15-DEC-1997  
LOCUS Borrelia burgdorferi (section 54 of 70) of the complete genome.  
ACCESSION AE001168 AE000783  
VERSION AE001168.1 GI:2688598  
KEYWORDS  
SOURCE Lyme disease spirochete.  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
REFERENCE 1 (bases 1 to 13271)

AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D.,  
Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S.,  
Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., Goceayne,J.,  
Venter,J.C., et al.  
TITLE Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi  
JOURNAL Nature 390 (6660), 580-586 (1997)  
MEDLINE 98065943  
REFERENCE 2 (bases 1 to 13271)  
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,  
Goceayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,  
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
Source  
Location/Qualifiers  
1. 13271  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
gene  
66. 1100  
/gene="BB0668"  
66. 1100  
/gene="BB0668"  
/note="similar to GP:1575447 percent identity: 98.83;  
identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="flagellar filament outer layer protein (flaA)"  
/protein\_id="AAC67025.1"  
/db\_xref="GI:2688608"  
/translation="MYMKRRKASILFFLSTVLEPAETDGLAEGSKRAEPGEVLDF  
AEIARDPSTRDLTNYVYVYSGASGVKPEDVVDGIMNWSYLLPSPARLQAYVYK  
NSVAVAVKSESKRYAGDTLLGVLEFSPYSOSAMIMPEFKIPYEGESGNOFLGKI  
GLDINKIKRIKYSVSLGVEIDLEVLFEEDMGMEVYSGMTLFGKGMADLIMSNPN  
YIPNISRIKDDVPNPLASSKMRKAFRYSKSHSSKSKSEQNFITYVDLRYLVDKLSY  
SDIDSESVFKYETSGETSLRKLKAHETFKYKLREKISMPESGFQNFVERKTES  
EKPESSPKN"  
gene  
1168. 3762  
/gene="BB0669"  
1168. 3762  
/gene="BB0669"  
/note="similar to PID:1113815 SP:Q44737 GB:AE000783  
percent identity: 99.31; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="chemotaxis histidine kinase (cheA-2)"  
/protein\_id="AAC67024.1"  
/db\_xref="GI:2688607"  
/translation="WEILDLENEELLGVFEPAQNLVDILEENISLEDPPNNSDTID  
EIPRAHRLKGSASISDMMELSDPHTIYEDVDAIRDKGVINNDVLLSSLDVVK  
EMALRIDGKYINDISDLKSKIKQFLYIDQTFIKRPGDSIKNNKINRDLSDSEIR  
EGLTGQVLRISVVFNSNSSEVENSGSLKFLNLIKNGSVLHTTPKEQJTEDEFLK  
RVDYLLYSIDIEGKSLSDSLNLESYLVDFENVEELKRLADEIKVDYDNEVNL  
DNFDTDEIDSLDLEVENOKLFYRLDFVDNDPMATISGLQMLAKSLKIRFSIP  
DSSELLADKPEDFYVYVYISNTSEESIAKKINLPDVYSPFEIKVNLSELSKRSKED  
DEAPREKNNKIKNSPTSVNLIRIDSKRIDVILINVSFAVISKSYNINSEMTLLFY  
NPNFYDQESFORNFLDKIVREDAKLTLEDELSINLSKFMKAKAKDISELR  
NSFFRLQNFRTSGRLSRIITDLSHESYLKTRMLPISNIFSRFTRVADLSKRLKIV  
NLKMEGETLELDKSYIDLVDPRLMKCVNSMDHGLYEEVEVRKGSAGCTIILRAKN  
EGNYSIIEEDGIGIDPKVIRRLKIEGTIKEDAIYSDPELINLIFAPGFTAVQV  
DLSGRGVGLDVVKKSIKRLNGTILVESIGTIFPKIKPLTVLIQGLVKSGETY  
VILPNLVETFRIRIETHDKILENHEVNLDEVTYVLRKLKINPFDSDSLIEKELI  
VYNTSNMKAIYVDSIIGEDFPVVKPIKDKSSAGIYGATITLNGKVVLLIIDVFKFL  
DLQDKTKE"  
gene  
3769. 5169

gene  
CDS  
/gene="BB0670"  
3769..5169  
/note="similar to GP:1881576 percent identity: 98.93;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="purine-binding chemotaxis protein (chew-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="GI:2688606"  
/translation="MQKEIYFGRKILDDKNSKLTJNDFKVSFELGSDHYLVDM  
QVKEIRKSNFVYNNAKYVAGLDNLRGEIIPIDLRIMPLEFNKXIDEDIMVAKN  
EDLIGVYDKINNVESIDSLIDPHVPLSODSLINIKGVADNEKYLILANFKI  
FNVEGEKILKPGONFVERSDPVDNDLLENSKNDSPKSSKNSFNNSLTNN  
TRAVNENIKKRLKLYSRNASLVNDFLEKVCYKFMVDTDLPCDSFLNERYSKSSG  
NMGADCLSEERNEIYESRLSPNNLNSIFNVLEIGCGSGKETMALANLSEYVAF  
KLTALDNDLKVETSRVSESEIGSEIYSRNSQSGVYKERSSELSNVEYS  
DALFSDLPDLNGLVFLKDVLCFLDSKDDILNITASKAIKGLVLVGLGNEELKNNDV  
FIKERTKYFNLYRKV"  
5181..5666  
/gene="BB0671"  
5181..5666  
/gene="BB0671"  
/note="similar to GP:1408273 percent identity: 98.76;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (chex)"  
/protein\_id="AAC67022.1"  
/db\_xref="GI:2688605"  
/translation="MRIDYIEPFLDAASSVLRLDMLVENIEMKPGKLSINOKIGVS  
VIVGLAGSEGIITIDMDIETFLFVASKINEFEYDFDEDEKEMVAALTEKGNITA  
GNFVTLHAKGVFDITPPAFIYGENMKISNGSEALIVFSLPDGKIIEVNAIRER  
V"  
5706..6146  
/gene="BB0672"  
5706..6146  
/gene="BB0672"  
/note="similar to GP:1408274 percent identity: 97.95;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis response regulator (chex-3)"  
/protein\_id="AAC67021.1"  
/db\_xref="GI:2688604"  
/translation="MIQKTTIADSSSRKRGINYDTGIPNVLIYDVSFTVKOLTOI  
FTSEGFNIIDTADGEAVIKYKNHPNIDITLDTMKMGITCLSNIMEFDKNAR  
VIMISALGKQELVNDCLIKGAKTFIVKPLDRAKVLQRVKSVFVK"  
complement(6187..6702)  
/gene="BB0673"  
complement(6187..6702)  
/gene="BB0673"  
/note="similar to GP:1408275 percent identity: 97.56;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC67020.1"  
/db\_xref="GI:2688603"  
/translation="MKIQIILMLALDLPLNARLLDISIEKRADEIKYSSNYLIL  
EKEYTNPTSEIKNIYKLTTEHFWKSLNKTNSLNSNKKKAKYIIOSELIDDK  
PLKVKIIFKKNKNGIFKSHSLITYTKKPKLELYIENNAEPLKIFNLNITYELKNIDK  
ISNEMIFPERE"  
complement(6699..7745)  
/gene="BB0674"  
complement(6699..7745)  
/gene="BB0674"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0674"  
/protein\_id="AAC67027.1"

gene  
CDS  
/db\_xref="GI:2688610"  
/translation="MVKRSIFLKAIIIFISFELILELSIILFLPKIRPALIFLGF  
LFDYFIFELKIRKAVYSORLEIVVRNNLEDDITHCLIPAFSSVOLKNIIYAE  
TLNPLMLSEKRLRLRLRNDLIIEITYNSKEKRLIILARFRTSMALLPFTFFI  
IISSSKIVSIEPEKQENIKINISIIENAVIKERPFILIIKEKDDIYKSDSEIV  
YVSPSERVIMEKTEFYIDKYLORXSDDIIGFLTLPASPIFLPMNFKPKFASFL  
NPILMTKILQDPLEYRKIQIPFTLSEKVEYELAKSFNLLLEKINRSKRKIPLEI  
EKVKTIINKNOERK"  
complement(7739..8602)  
/gene="BB0675"  
complement(7739..8602)  
/gene="BB0675"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0675"  
/protein\_id="AAC67026.1"  
/db\_xref="GI:2688609"  
/translation="MNTKPYFPLVLYLFNHSIKSLASIEKEILYLNKKNKRTIA  
TIRKNDPSEIKDLIOYVKDKDIMTPPVLSIEADINIKYPLESKELTKNDMLI  
FNVKVNSSLRKEEFYFNFTISNGYTFYINKLFEKNSYTYILQKENKALYSDII  
KNYKILLKLVIVIKYCFEKGIELTKNIESTSKAISDTDFLDEKTAKLIESEFK  
YETLOTMSPISITLIAIFSARARPKYKNPVKGFISYDESMESIKQSGSREYDSRIIK  
ELSEIAKVNKM"  
complement(8599..9261)  
/gene="BB0676"  
complement(8599..9261)  
/gene="BB0676"  
/note="similar to GB:142023.SP:P44755.PID:1004013  
PID:1222502.PID:1204815 percent identity: 25.58;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoglycolate phosphatase (gph)"  
/protein\_id="AAC67019.1"  
/db\_xref="GI:2688602"  
/translation="MKIKACIPDMQDVLNSINDIAFNSNALSNGYSKIELSKENA  
LVGRGNKRVIDTLKLSLEHDPNLOEKLKYKEVYKNNLSFOKPPENIKPLET  
MKLNIPGILSKNHEELINLVKNFIEGILFEFVGSNFPKPKDPNALMILEL  
NAQKEIAYIGSDVDMLTALNAGFPIGVSWGFSVQELKESGAKHIIHNPDELIDL  
IK"  
9504..11114  
/gene="BB0677"

Query Match 100.0%; Score 21; DB 1; Length 13271;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atgaaagagaagctaaagt 21
Db	75	ATGAAAGGAAGCTAAAGT 95

RESULT 3

LEPOR	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
LEPOR	3885 bp	DNA	PLN	03-NOV-1994	LEPOR	esculentum polyubiquitin repeats.	Esculentum polyubiquitin repeats.
						direct repeat; inverted repeat; polyubiquitin gene; ubiquitin.	

REFERENCE 1 (bases 1 to 3885)  
Kollitnik, J.K. and Piltzner, U.M.  
Structure of a heptubiquitin gene from tomato  
Plant Physiol. 104 (1), 299-300 (1994)  
MEDLINE 94159799  
REFERENCE 2 (bases 1 to 3885)

**AUTHORS** Piltzner, U.M.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (03-JUN-1993) U.M. Piltzner, Botanisches Institut der  
LMU, Menzinger Str 67, 8000 Muenchen 19, FRG  
**FEATURES** Location/Qualifiers

Query Match	95.28;	Score 20;	DB 8;	Length 3885;
Best Local Similarity	100.0%;	Pred. No. 1.2e+02;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps
Oy	1 atgaaagaagctaagaag	20		

Db 78 ATGAAAGGAACCTAAAG 97

|||||

RESULT 4  
AC068117/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 183289)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 2, clone RP11-353K11  
Unpublished  
2 (bases 1 to 183289)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

DNA  
HTG  
map 2, WORKING DRAFT

AC068117 183289 bp  
Homo sapiens chromosome 2 clone RP11-353K11  
SEQUENCE, 32 unordered pieces.  
AC068117  
AC068117.2 GI:8140914  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L9905
Center clone name: 353_K11
-----
Summary Statistics
Sequencing vector: M13/ M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167319 bases at least Q40
Consensus quality: 174945 bases at least Q30
Consensus quality: 178209 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 180189; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```



\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1287: contig of 1287 bp in length  
1288 1387: gap of 100 bp  
1388 3151: contig of 1764 bp in length  
3152 3251: gap of 100 bp  
3252 4709: contig of 1458 bp in length  
4710 4809: gap of 100 bp  
4810 6242: contig of 1433 bp in length  
6243 6342: gap of 100 bp  
6343 8667: contig of 2325 bp in length  
8668 8767: gap of 100 bp  
8768 10535: contig of 1768 bp in length  
10536 10635: gap of 100 bp  
10636 12007: contig of 1372 bp in length  
12008 12107: gap of 100 bp  
12108 14409: contig of 2302 bp in length  
14410 14509: gap of 100 bp  
14510 16536: contig of 2027 bp in length  
16537 16636: gap of 100 bp  
16637 19520: contig of 2884 bp in length  
19521 19620: gap of 100 bp  
19621 22669: contig of 3049 bp in length  
22670 22769: gap of 100 bp  
22770 26259: contig of 3490 bp in length  
26260 26359: gap of 100 bp  
26360 29359: contig of 3000 bp in length  
29360 29459: gap of 100 bp  
29460 31820: contig of 2361 bp in length  
31821 31920: gap of 100 bp  
31921 34836: contig of 2916 bp in length  
34837 34936: gap of 100 bp  
34937 38183: contig of 3247 bp in length  
38184 38283: gap of 100 bp  
38284 40743: contig of 2460 bp in length  
40744 40843: gap of 100 bp  
40844 46140: contig of 5297 bp in length  
46141 46240: gap of 100 bp  
46241 52339: contig of 6099 bp in length  
52340 52439: gap of 100 bp  
52440 57689: contig of 5250 bp in length  
57690 57789: gap of 100 bp  
57790 65948: contig of 8159 bp in length  
65949 66048: gap of 100 bp  
66049 74025: contig of 7977 bp in length  
74026 74125: gap of 100 bp  
74126 80421: contig of 6296 bp in length  
80422 80521: gap of 100 bp  
80522 89984: contig of 9463 bp in length  
89985 90084: gap of 100 bp  
90085 99029: contig of 8945 bp in length  
99030 99129: gap of 100 bp  
99130 109926: contig of 10797 bp in length  
109927 110026: gap of 100 bp  
110027 120095: contig of 10069 bp in length  
120096 120195: gap of 100 bp  
120196 131804: contig of 11609 bp in length  
131805 131904: gap of 100 bp  
131905 142878: contig of 10974 bp in length  
142879 142978: gap of 100 bp  
142979 153656: contig of 10678 bp in length  
153657 153756: gap of 100 bp  
153757 169035: contig of 15279 bp in length  
169036 169135: gap of 100 bp  
169136 183289: contig of 14154 bp in length.

FEATURES  
Source

1.183289  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"

/clone="RP11-353K11"  
/clone\_1lb="RP11 Human Male BAC"  
1.1287  
/note="assembly\_fragment"  
misc\_feature  
1388.3151  
/note="assembly\_fragment"  
misc\_feature  
3252.4709  
/note="assembly\_fragment"  
misc\_feature  
4810.6242  
/note="assembly\_fragment"  
misc\_feature  
6343.8667  
/note="assembly\_fragment"  
misc\_feature  
8768.10535  
/note="assembly\_fragment"  
misc\_feature  
10636.12007  
/note="assembly\_fragment"  
misc\_feature  
12108.14409  
/note="assembly\_fragment"  
misc\_feature  
14510.16536  
/note="assembly\_fragment"  
misc\_feature  
16637.19520  
/note="assembly\_fragment"  
misc\_feature  
19621.22669  
/note="assembly\_fragment"  
misc\_feature  
22770.26259  
/note="assembly\_fragment"  
misc\_feature  
26360.29359  
/note="assembly\_fragment"  
misc\_feature  
29460.31820  
/note="assembly\_fragment"  
misc\_feature  
31921.34836  
/note="assembly\_fragment"  
misc\_feature  
34937.38183  
/note="assembly\_fragment"  
misc\_feature  
38284.40743  
/note="assembly\_fragment"  
misc\_feature  
vector\_side:right  
40844.46140  
/note="assembly\_fragment"  
misc\_feature  
46241.52339  
/note="assembly\_fragment"  
misc\_feature  
52440.57689  
/note="assembly\_fragment"  
misc\_feature  
57790.65948  
/note="assembly\_fragment"  
misc\_feature  
66049.74025  
/note="assembly\_fragment"  
misc\_feature  
74126.80421  
/note="assembly\_fragment"  
misc\_feature  
80522.89984  
/note="assembly\_fragment"  
misc\_feature  
90085.99029  
/note="assembly\_fragment"  
misc\_feature  
99130.109926  
/note="assembly\_fragment"  
misc\_feature  
110027.120095  
/note="assembly\_fragment"  
misc\_feature  
120196.131804  
/note="assembly\_fragment"

Query Match 92.4% Score 19.4: DB 2: Length 183289;  
Best Local Similarity 95.2% Pred. No. 2.4e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaagaagaagctaaagt 21  
|||||  
Db 52810 ATGAAAGGAACATAAACT 52790  
|||||  
RESULT 5  
AC068539 AC068539 198207 bp DNA HTG 12-AUG-2000  
LOCUS

DEFINITION Homo sapiens chromosome 2 clone RP11-145M22, WORKING DRAFT  
 SEQUENCE, 13 unordered pieces.  
 ACCESSION AC068539  
 VERSION AC068539.5 GI:9719795  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 198207)  
 Waterston, R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 198207)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (03-MAY-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Aug 7, 2000 this sequence version replaced g1:8571789.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H.NH0145M22  
 ----- Summary Statistics -----  
 Sequencing vector: M13, 100%  
 Sequencing vector: Plasmid: 0%  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 190362 bases at least Q40  
 Consensus quality: 194318 bases at least Q30  
 Insert size: 20000; agarose-fp  
 Insert size: 197007; sum-of-ctrls  
 Quality coverage: 4.96 in Q20 bases; agarose-fp  
 Quality coverage: 5.07 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 \* 2651: contig of 2651 bp in length  
 \* 2652: gap of unknown length  
 \* 2752: contig of 2781 bp in length  
 \* 5532: gap of unknown length  
 \* 5533: contig of 5703 bp in length  
 \* 11335: gap of unknown length  
 \* 11336: contig of 5703 bp in length  
 \* 11436: gap of unknown length  
 \* 18140: contig of 6705 bp in length  
 \* 18240: gap of unknown length  
 \* 18241: contig of 7359 bp in length  
 \* 25599: gap of unknown length  
 \* 25600: contig of unknown length  
 \* 25700: gap of unknown length  
 \* 36129: contig of 10429 bp in length  
 \* 36228: gap of unknown length  
 \* 36229: contig of 12825 bp in length  
 \* 49053: gap of unknown length  
 \* 49154: contig of 12234 bp in length  
 \* 49154: gap of unknown length  
 \* 61388: contig of unknown length  
 \* 61488: gap of unknown length  
 \* 74323: contig of 12836 bp in length  
 \* 74324: gap of unknown length  
 \* 74324: contig of 14470 bp in length  
 \* 88894: gap of unknown length  
 \* 88994: contig of 19852 bp in length  
 \* 108845: gap of unknown length  
 \* 108846: contig of 34470 bp in length  
 \* 108946: gap of unknown length  
 \* 143416: gap of unknown length

FEATURES \* 143516 198207: contig of 54692 bp in length.  
 source 1..198207  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /clone="RP11-145M22"

misc\_feature  
 /note="assembly\_name:Contig19"  
 2752..5532  
 /note="assembly\_name:Contig20"  
 5633..11335  
 /note="assembly\_name:Contig21"  
 11436..18140  
 /note="assembly\_name:Contig22"  
 18241..25599  
 /note="assembly\_name:Contig23"  
 25700..36128  
 /note="assembly\_name:Contig24"  
 36229..49053  
 /note="assembly\_name:Contig25"  
 49154..61387  
 /note="assembly\_name:Contig26"  
 clone\_end:sp6  
 vector\_side:right"  
 61488..74323  
 /note="assembly\_name:Contig27"  
 74424..88893  
 /note="assembly\_name:Contig28"  
 88994..108845  
 /note="assembly\_name:Contig29"  
 clone\_end:t7  
 vector\_side:left"  
 108946..143415  
 /note="assembly\_name:Contig30"  
 143516..198207  
 /note="assembly\_name:Contig31"  
 198207..54692  
 /note="assembly\_name:Contig32"  
 54692..54692

Query Match 92.4%; Score 19.4; DB 2; Length 198207;  
 Best Local Similarity 95.2%; Pred. No. 2.4e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 atgaaaggaagcraaagt 21  
 |||||||||||  
 Db 103013 ATGAAAGGAAACTTAAAGT 103033

RESULT 6  
 CEM25C12/3 33045 bp DNA INV 20-JUN-2001  
 LOCUS CEM25C12/3 33045 bp DNA INV 20-JUN-2001  
 DEFINITION Caenorhabditis elegans cosmid T25C12, complete sequence.  
 ACCESSION Z66566  
 VERSION Z66566.1 GI:1051339  
 HTG: jln-14.  
 KEYWORDS Caenorhabditis elegans.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;  
 Rhabditiida; Rhabditiida; Peloderiinae; Caenorhabditis.  
 1 (bases 1 to 33045)  
 TITLE none.  
 AUTHORS none.  
 TITLE Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium  
 Science 282 (5396), 2012-2018 (1998)  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 99069613  
 REMARK The C. elegans Sequencing Consortium.  
 REFERENCE 2 (bases 1 to 33045)  
 AUTHORS Kershaw, J.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-NOV-1995) Nematode Sequencing Project, Sanger Centre,  
 Hinxton, Cambridge CB10 1RO, England and Department of Genetics,



comes from this gene  
 cDNA EST yk22a7.3 comes from this gene; cDNA EST yk26b2.3  
 comes from this gene  
 cDNA EST yk38a1.3 comes from this gene; cDNA EST yk42e7.3  
 comes from this gene  
 cDNA EST yk13d10.5 comes from this gene; cDNA EST yk2h10.5  
 comes from this gene  
 cDNA EST yk4b3.5 comes from this gene; cDNA EST yk5f7.5  
 comes from this gene  
 cDNA EST yk7a3.5 comes from this gene; cDNA EST yk13c12.5  
 comes from this gene  
 cDNA EST yk10b4.5 comes from this gene; cDNA EST yk23b5.5  
 comes from this gene  
 cDNA EST yk26b2.5 comes from this gene; cDNA EST yk38a1.5  
 comes from this gene  
 cDNA EST yk22a7.5 comes from this gene; cDNA EST  
 yk38b10.3 comes from this gene  
 cDNA EST yk38b10.5 comes from this gene; cDNA EST  
 yk259f5.3 comes from this gene  
 cDNA EST yk259f5.5 comes from this gene; cDNA EST  
 yk295a1.3 comes from this gene  
 cDNA EST yk295a1.5 comes from this gene; cDNA EST  
 yk294g11.3 comes from this gene  
 cDNA EST yk294g11.5 comes from this gene; cDNA EST  
 yk285b2.3 comes from this gene  
 cDNA EST yk285b2.5 comes from this gene; cDNA EST  
 yk276d2.3 comes from this gene  
 cDNA EST yk276d2.5 comes from this gene; cDNA EST  
 yk243d6.3 comes from this gene  
 cDNA EST yk243d6.5 comes from this gene; cDNA EST  
 yk239d3.3 comes from this gene  
 cDNA EST yk239d3.5 comes from this gene; cDNA EST  
 yk226g10.3 comes from this gene

Query Match 87.6%; Score 18.4; DB 3; Length 33045;  
 Best Local Similarity 95.0%; Pred. No. 6; le+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaag 20  
 ||||| ||||| ||||| |||||  
 DB 1344 ATGAAAGAAACTRAAAG 1325

RESULT 7  
 AC091103 62555 bp DNA HTG 29-MAR-2001  
 LOCUS Homo sapiens chromosome 18 clone RP11-749G1 map 18; LOW-PASS  
 DEFINITION  
 AC091103 SEQUENCE SAMPLING.  
 AC091103.1 GI:13488002  
 VERSION HTG; HTGS\_PHASE0.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 62555)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 18, clone RP11-749G1  
 Unpublished  
 2 (bases 1 to 62555)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
 Batra, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,  
 Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M.,  
 Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K.,  
 Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,  
 Iliev, I., Johnson, R., Jones, C., Karatas, A., Labroque, K.,  
 Iamazaes, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 Maclean, C., Macdonald, P., Marcis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Menes, L.,  
 Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nordu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

# TITLE JOURNAL COMMENT

Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Roselli, M., Roy, A., Santos, R., Schaner, S., Schnupp, R., Seaman, S.,  
 Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
 Theodore, J., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: 749\_G1  
 -----  
 \* NOTE: This record contains 70 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1 795: contig of 795 bp in length  
 \* 796 895: gap of 100 bp  
 \* 896 1636: contig of 741 bp in length  
 \* 1637 1736: gap of 100 bp  
 \* 1737 2537: contig of 801 bp in length  
 \* 2538 2637: gap of 100 bp  
 \* 2638 3408: contig of 771 bp in length  
 \* 3409 3508: gap of 100 bp  
 \* 3509 4313: contig of 805 bp in length  
 \* 4314 4413: gap of 100 bp  
 \* 4414 5220: contig of 807 bp in length  
 \* 5221 5320: gap of 100 bp  
 \* 5321 6125: contig of 805 bp in length  
 \* 6126 6225: gap of 100 bp  
 \* 6226 6993: contig of 768 bp in length  
 \* 6994 7093: gap of 100 bp  
 \* 7094 7891: contig of 798 bp in length  
 \* 7892 7991: gap of 100 bp  
 \* 7992 8783: contig of 792 bp in length  
 \* 8784 8883: gap of 100 bp  
 \* 8884 9674: contig of 791 bp in length  
 \* 9675 9774: gap of 100 bp  
 \* 9775 10576: contig of 802 bp in length  
 \* 10577 10676: gap of 100 bp  
 \* 10677 11486: contig of 810 bp in length  
 \* 11487 11586: gap of 100 bp  
 \* 11587 12384: contig of 798 bp in length  
 \* 12385 12484: gap of 100 bp  
 \* 12485 13265: contig of 781 bp in length  
 \* 13266 13365: gap of 100 bp  
 \* 13366 14161: contig of 796 bp in length  
 \* 14162 14261: gap of 100 bp  
 \* 14262 15053: contig of 792 bp in length  
 \* 15054 15153: gap of 100 bp  
 \* 15154 15955: contig of 802 bp in length  
 \* 15956 16055: gap of 100 bp  
 \* 16056 16864: contig of 809 bp in length  
 \* 16865 16964: gap of 100 bp  
 \* 16965 17773: contig of 809 bp in length

1774 17873: gap of 100 bp  
17874 18705: contig of 832 bp in length  
18706 18805: gap of 100 bp  
18806 18629: contig of 824 bp in length  
19630 19729: gap of 100 bp  
19730 20538: contig of 809 bp in length  
20539 20638: gap of 100 bp  
20639 21461: contig of 823 bp in length  
21462 21561: gap of 100 bp  
21562 22380: contig of 819 bp in length  
22381 22480: gap of 100 bp  
22481 23278: contig of 798 bp in length  
23279 23378: gap of 100 bp  
23379 24169: contig of 791 bp in length  
24170 24269: gap of 100 bp  
24270 25056: contig of 787 bp in length  
25057 25156: gap of 100 bp  
25157 25971: contig of 815 bp in length  
25972 26071: gap of 100 bp  
26072 26864: contig of 793 bp in length  
26865 26964: gap of 100 bp  
26965 27778: contig of 814 bp in length  
27779 27878: gap of 100 bp  
27879 28661: contig of 783 bp in length  
28662 28761: gap of 100 bp  
28762 29545: contig of 784 bp in length  
29546 29645: gap of 100 bp  
29646 30418: contig of 773 bp in length  
30419 30518: gap of 100 bp  
30519 31296: contig of 778 bp in length  
31297 31396: gap of 100 bp  
31397 32204: contig of 808 bp in length  
32205 32304: gap of 100 bp  
32306 33095: contig of 791 bp in length  
33096 33195: gap of 100 bp  
33196 33982: contig of 787 bp in length  
33983 34082: gap of 100 bp  
34083 34886: contig of 804 bp in length  
34887 34986: gap of 100 bp  
34987 35799: contig of 813 bp in length  
35800 35899: gap of 100 bp  
35900 36689: contig of 790 bp in length  
36690 36789: gap of 100 bp  
36790 37593: contig of 804 bp in length  
37594 37693: gap of 100 bp  
37694 38486: contig of 793 bp in length  
38487 38586: gap of 100 bp  
38587 39386: contig of 800 bp in length  
39387 39486: gap of 100 bp  
39487 40281: contig of 795 bp in length  
40282 40381: gap of 100 bp  
40382 41179: contig of 798 bp in length  
41180 41279: gap of 100 bp  
41280 42094: contig of 815 bp in length  
42095 42194: gap of 100 bp  
42195 42992: contig of 798 bp in length  
42993 43092: gap of 100 bp  
43093 43887: contig of 795 bp in length  
43888 43987: gap of 100 bp  
43988 44752: contig of 765 bp in length  
44753 44852: gap of 100 bp  
44853 45626: contig of 774 bp in length  
45627 45726: gap of 100 bp  
45727 46507: contig of 781 bp in length  
46508 46607: gap of 100 bp  
46609 47403: contig of 796 bp in length  
47404 47503: gap of 100 bp  
47504 48253: contig of 750 bp in length  
48254 48353: gap of 100 bp  
48354 49177: contig of 824 bp in length  
49178 49277: gap of 100 bp  
49278 50094: contig of 817 bp in length  
50095 50194: gap of 100 bp

50195 50959: contig of 765 bp in length  
50960 51059: gap of 100 bp  
51060 51849: contig of 790 bp in length  
51850 51949: gap of 100 bp  
51950 52741: contig of 792 bp in length  
52742 52841: gap of 100 bp  
52842 53635: contig of 794 bp in length  
53636 53735: gap of 100 bp  
53736 54476: contig of 741 bp in length  
54477 54576: gap of 100 bp  
54577 55375: contig of 799 bp in length  
55378 55475: gap of 100 bp  
55476 56258: contig of 783 bp in length  
56259 56358: gap of 100 bp  
56359 57161: contig of 803 bp in length  
57162 57261: gap of 100 bp  
57262 58053: contig of 792 bp in length  
58054 58153: gap of 100 bp  
58154 58959: contig of 806 bp in length  
58960 59059: gap of 100 bp  
59060 59865: contig of 806 bp in length  
59866 59965: gap of 100 bp  
59966 60765: contig of 800 bp in length  
60766 60865: gap of 100 bp  
60866 61660: contig of 795 bp in length  
61661 61760: gap of 100 bp

Query Match 87.6%; Score 18.4; DB 2; Length 62555;  
Best Local Similarity 95.0%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaagaagaaagctaaagt 21  
Db 46462 TGAAGAACAAAGCTAAAGT 46481  
||||||| |||||||||

RESULT 8  
AC073440/C  
LOCUS  
DEFINITION  
AC073440 Homo sapiens chromosome 12 clone RP11-100E14, WORKING DRAFT  
SEQUENCE, 7 unordered pieces.  
AC073440 AC073440.10 GI:9972211  
VERSION  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 139064)  
Muzny,D.M., Adams,C., Bailey,M., Barbarta,J., Blankenburg,K.,  
Bodola,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,  
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,  
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,  
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,  
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorell,J.H., Gorell,L.L.,  
Guevara,W., Harris,K., Hernandez,U., Hodgson,A., Hugues,M.,  
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,  
Kelly,S., Kondejowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,  
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,  
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,  
Morris,S., Naeh,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,  
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,  
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,  
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R.,  
Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,  
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,  
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and  
Gibbs,R.  
Direct Submission  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 139064)  
Worley,K.C.  
AUTHORS  
TITLE  
Direct Submission

JOURNAL Submitted (18-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 6, 2000 this sequence version replaced gi:9964656.

## ----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

## ----- Project Information

Center project name: HBCF

Center clone name: RP11-100E14

## ----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 75% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 131675 bases at least Q40

Consensus quality: 135730 bases at least Q30

Consensus quality: 138146 bases at least Q20

Estimated insert size: 134201; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see <http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html>).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 54266: contig of 54266 bp in length  
\* 54367 54366: gap of unknown length  
\* 102107 102107: contig of 47741 bp in length  
\* 102208 102207: gap of unknown length  
\* 132268 132268: contig of 30061 bp in length  
\* 132369 132368: gap of unknown length  
\* 134381 134381: contig of 2013 bp in length  
\* 134382 134481: gap of unknown length  
\* 134482 136148: contig of 1668 bp in length  
\* 136150 136249: gap of unknown length  
\* 136250 137922: contig of 1673 bp in length  
\* 137923 138022: gap of unknown length  
\* 138023 139064: contig of 1042 bp in length.

## FEATURES

Location/Qualifiers

1..139064

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="RP11-100E14"

BASE COUNT 43158 a 27043 c 26771 g 41475 t 617 others  
ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 139064;  
Best Local Similarity 95.0%; Pred. No. 6.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaagaagaaagctaaagt 21  
||||| |||||||

Db 4584 TGAAGAAGCAAGCTAAAGT 4565

RESULT 9  
AC011421 148142 bp DNA PRI 18-APR-2000  
LOCUS Homo sapiens chromosome 5 clone CTD-230305, complete sequence.  
AC011421  
AC011421.2 GI:6102634  
VERSION HTG.  
KEYWORDS

## SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 148142)  
DOE Joint Genome Institute and Stanford Human Genome Center.

## AUTHORS

Unpublished

## REFERENCE

2 (bases 1 to 148142)  
DOE Joint Genome Institute.

## AUTHORS

Direct Submission  
Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## TITLE

DOE Joint Genome Center and Stanford Human Genome Center.

## REFERENCE

Submitted (23-OCT-1999) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

## AUTHORS

4 (bases 1 to 148142)  
DOE Joint Genome Institute and Stanford Human Genome Center.

## TITLE

Direct Submission  
Submitted (01-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

## REFERENCE

5 (bases 1 to 148142)  
DOE Joint Genome Institute and Stanford Human Genome Center.

## AUTHORS

Direct Submission  
Submitted (02-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

## REFERENCE

6 (bases 1 to 148142)  
DOE Joint Genome Institute and Stanford Human Genome Center.

## AUTHORS

Direct Submission  
Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

## TITLE

On Oct 23, 1999 this sequence version replaced gi:6013521.

## COMMENT

Draft Sequence Produced by DOE Joint Genome Institute

[www.jgi.doe.gov](http://www.jgi.doe.gov)

Finishing completed at Stanford Human Genome Center

[www.shgc.stanford.edu](http://www.shgc.stanford.edu)

Quality: Phrap Quality >=40 99.9% of Sequence;

Estimated Total Number of Errors is 0.1.

STS Content:

WT-6744 G05721

WT-6988 G06342.

## FEATURES

Location/Qualifiers

1..148142

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CTD-230305"

BASE COUNT 50576 a 26865 c 24848 g 45853 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 148142;  
Best Local Similarity 95.0%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaagaagaaagctaaag 20  
||||| |||||||

Db 40106 ATGAAGAAGCAAGCTAAAG 40125

RESULT 10  
AC091111 156048 bp DNA HTG 11-AUG-2001  
LOCUS Homo sapiens chromosome 18 clone RP11-837W2, map 18, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*; 1 ordered pieces.  
AC091111  
AC091111.3 GI:15028541  
VERSION HTG; HTGS\_PHASE2; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
KEYWORDS human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 156048)
TITLE	2 (bases 1 to 156048)
JOURNAL	Unpublished
REFERENCE	Homo sapiens chromosome 18, clone RP11-837M2
AUTHORS	2 (bases 1 to 156048)
	Barn, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Barn, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heeford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kartas, A., Larocque, K., Lamazeres, R., Landers, T., Lehotzky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McGowan, P., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Mihova, T., Miengna, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, R., Raymond, C., Reiter, R., Ribback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Soungue, C., Spencer, C., Stange-Thomas, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travers, T., Trifillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (23-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Jul 29, 2001 this sequence version replaced gi:14389354.
COMMENT	All repeats were identified using RepeatMasker: Smt, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
	----- Genome Center
	Center: Whitehead Institute/ MIT Center for Genome Research
	Web site: http://www-seq.wi.mit.edu
	Contact: sequence_submissions@genome.wi.mit.edu
	----- Project Information
	Center project name: L12690
	Center clone name: 837_M.2
	-----
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 1 contigs. Gaps between the contigs
	* are represented as runs of N. The order of the pieces
	* is believed to be correct as given, however the sizes
	* of the gaps between them are based on estimates that have
	* provided by the submitter.
	* This sequence will be replaced
	* by the finished sequence as soon as it is available and
	* the accession number will be preserved.
	* 1 156048: contig of 156048 bp in length.
	Location/Qualifiers
FEATURES	1. 156048
source	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="18"
	/map="18"
	/clone="RP11-837M2"
	/clone_11b="RPC1-11 Human Male BAC"
BASE COUNT	46624 a 30207 c 30590 g 48568 t 59 others
ORIGIN	
Query Match	87.6%, Score 18.4; DB 2; Length 156048;
Best Local Similarity	95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
2 tgaagaagaagctataaagt 21	
Db 125436 TGAAGAAGCAAGCTATAAAGT 125455	

RESULT	11
AC023145	
LOCUS	AC023145 159012 bp DNA HTG 07-JUL-2000
DEFINITION	Homo sapiens chromosome 8 clone RP11-621J21, WORKING DRAFT
ACCESSION	AC023145
VERSION	AC023145.4 GI:7344691
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 159012)
TITLE	Waterston, R.H.
JOURNAL	The sequence of Homo sapiens clone
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 159012)
TITLE	Waterston, R.H.
JOURNAL	Direct Submission
COMMENT	Submitted (08-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Mar 30, 2000 this sequence version replaced gi:7253502.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0621J21
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid: 0%
Chemistry: Dye-Primer ET; 91% of reads
Chemistry: Dye-terminator Big Dye; 9% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 149846 bases at least Q40
Consensus quality: 151997 bases at least Q30
Consensus quality: 153141 bases at least Q20
Insert size: 184000; agarose-fp
Insert size: 156512; sum-of-contigs
Quality coverage: 3.44 in Q20 bases; agarose-fp
Quality coverage: 4.11 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1282: contig of 1282 bp in length
*
1283      1382: gap of unknown length
*
1383      2732: contig of 1350 bp in length
*
2733      2832: gap of unknown length
*
2833      4155: contig of 1323 bp in length
*
4156      4255: gap of unknown length
*
4256      6153: contig of 1900 bp in length
*
6156      6255: gap of unknown length
*
6256      7748: contig of 1493 bp in length
*
7749      7848: gap of unknown length
*
7849      10056: contig of 2208 bp in length
*
10057      10156: gap of unknown length
*
10157      12429: contig of 2273 bp in length
*
12430      12529: gap of unknown length
*
12530      15456: contig of 2927 bp in length
*
15457      15556: gap of unknown length
*
15557      20085: contig of 4529 bp in length
*
20086      20185: gap of unknown length
*
20186      23226: contig of 3041 bp in length
*
23227      23226: gap of unknown length

```

```
* 23327 26489: contig of 3163 bp in length
* 26580 26589: gap of unknown length
* 29068 29067: contig of 2478 bp in length
* 29168 29167: gap of unknown length
* 33180 33179: contig of 4012 bp in length
* 33280 33279: gap of unknown length
* 37504 37503: contig of 4224 bp in length
* 37604 37603: gap of unknown length
* 43401 43400: contig of 5797 bp in length
* 43501 43500: gap of unknown length
* 48719 48718: contig of 5219 bp in length
* 48820 48819: gap of unknown length
* 54617 54616: contig of 5797 bp in length
* 54717 54716: gap of unknown length
* 60644 60643: contig of 5927 bp in length
* 60744 60743: gap of unknown length
* 69313 69312: contig of 8569 bp in length
* 69413 69412: gap of unknown length
* 81229 81228: contig of 11816 bp in length
* 81329 81328: gap of unknown length
* 91246 91245: contig of 9917 bp in length
* 91346 91345: gap of unknown length
* 104493 104492: contig of 13147 bp in length
* 104593 104592: gap of unknown length
* 116998 116997: contig of 12405 bp in length
* 117098 117097: gap of unknown length
* 127814 127813: contig of 10716 bp in length
* 127914 127913: gap of unknown length
* 141385 141384: contig of 13471 bp in length
* 141485 159012: contig of 17528 bp in length.
Location/Qualifiers
1. 159012
/organism="Homo sapiens"
/ld_xref="taxon:9606"
/chromosome="8"
/clone="RP11-621J21"
```

## FEATURES

```
misc_feature
/note="assembly_name:Contig2"
1383..2732
misc_feature
/note="assembly_name:Contig6"
2833..4155
misc_feature
/note="assembly_name:Contig7"
4256..6155
misc_feature
/note="assembly_name:Contig8"
6256..7748
misc_feature
/note="assembly_name:Contig9"
7849..10056
misc_feature
/note="assembly_name:Contig10"
10157..12429
misc_feature
/note="assembly_name:Contig11"
12530..15456
misc_feature
/note="assembly_name:Contig12"
15557..20085
misc_feature
/note="assembly_name:Contig13"
20186..23226
misc_feature
/note="assembly_name:Contig14"
clone_end:T7
vector_side:left"
23327..26489
misc_feature
/note="assembly_name:Contig15"
26590..29067
misc_feature
/note="assembly_name:Contig16"
29168..33179
misc_feature
/note="assembly_name:Contig17"
33280..37503
misc_feature
/note="assembly_name:Contig18"
37604..43400
misc_feature
/note="assembly_name:Contig19"
43501..48719
misc_feature
/note="assembly_name:Contig20"
48820..54616
misc_feature
/note="assembly_name:Contig21"
```

```
misc_feature 54717..60643
/note="assembly_name:Contig22"
misc_feature 60744..69312
/note="assembly_name:Contig23"
misc_feature 69413..81228
/note="assembly_name:Contig24"
clone_end:SP6
vector_side:left"
misc_feature 81329..91245
/note="assembly_name:Contig25"
misc_feature 91346..104492
/note="assembly_name:Contig26"
misc_feature 104593..116997
/note="assembly_name:Contig27"
misc_feature 117098..127813
/note="assembly_name:Contig28"
misc_feature 127914..141384
/note="assembly_name:Contig29"
141485..159012
/note="assembly_name:Contig30"
BASE COUNT 53129 a 27457 c 27795 g 48121 t 2510 others
ORIGIN
```

```
Query Match 87.6%; Score 18.4; DB 2; Length 159012;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Caps 0;
OY 2 tgaagaagaaagctaaagt 21
|||||
Db 43587 TGAAGAAGCAATCTAAAGT 43606
```

```
RESULT 12
AP001493/c
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-779H19 map 18q21, WORKING
DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION
AP001493.2 GI:8117351
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT
KEYWORDS
Homo sapiens DNA, clone:RP11-779H19.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 161624)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 161,624 genomic DNA of 18q21
Published Only in Database (2000) In press
2 (bases 1 to 161624)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@sc.riken.go.jp,
URL:http://npg.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7288187.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://npg.gsc.riken.go.jp/
Contact: hattori@sc.riken.go.jp
----- Project Information
Center project name: Humdrat18
Center clone name: RP11-779H19
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
```

## COMMENT



Consensus quality: 150399 bases at least Q40  
Consensus quality: 155891 bases at least Q30  
Consensus quality: 158205 bases at least Q20  
Insert size: 159424; sum-of-ctrls  
Quality coverage: 5.50x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 32754 contig of 32754 bp in length  
32855 48719 contig of 15865 bp in length  
48820 64088 contig of 15269 bp in length  
64189 76256 contig of 12068 bp in length  
76357 86295 contig of 9939 bp in length  
86396 94552 contig of 8157 bp in length  
94653 102595 contig of 7943 bp in length  
102596 110667 contig of 7972 bp in length  
110768 116555 contig of 5788 bp in length  
116656 122099 contig of 5444 bp in length  
122200 125461 contig of 3262 bp in length  
125562 131077 contig of 5516 bp in length  
131178 135924 contig of 4747 bp in length  
136025 138309 contig of 2285 bp in length  
138310 138409 contig of 1082 bp in length

Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 32754 contig of 32754 bp in length  
32755 32854: gap of 100 bp  
32855 48719: contig of 15865 bp in length  
48720 48819: gap of 100 bp  
48820 64088: contig of 15269 bp in length  
64089 64188: gap of 100 bp  
64189 76256: contig of 12068 bp in length  
76257 76356: gap of 100 bp  
76357 86295: contig of 9939 bp in length  
86296 86395: gap of 100 bp  
86396 94552: contig of 8157 bp in length  
94553 102595: gap of 100 bp  
94553 102595: contig of 7943 bp in length  
102596 102695: gap of 100 bp  
102696 110667: contig of 7972 bp in length  
110668 110767: gap of 100 bp  
110768 116555: contig of 5788 bp in length  
116556 116655: gap of 100 bp  
116656 122099: contig of 5444 bp in length  
122100 122199: gap of 100 bp  
122200 125461: contig of 3262 bp in length  
125462 125561: gap of 100 bp  
125562 131077: contig of 5516 bp in length  
131078 131177: gap of 100 bp  
131178 135924: contig of 4747 bp in length  
135925 136024: gap of 100 bp  
136025 138309: contig of 2285 bp in length  
138310 138409: gap of 100 bp

\* 138410 140832: contig of 2423 bp in length  
\* 140833 140932: gap of 100 bp  
\* 140933 144613: contig of 3681 bp in length  
\* 144614 144713: gap of 100 bp  
\* 144714 147380: contig of 2667 bp in length  
\* 147381 147480: gap of 100 bp  
\* 147481 150820: contig of 3340 bp in length  
\* 150821 150920: gap of 100 bp  
\* 150921 154114: contig of 3194 bp in length  
\* 154115 154214: gap of 100 bp  
\* 154215 156288: contig of 2074 bp in length  
\* 156289 156388: gap of 100 bp  
\* 156389 158482: contig of 2094 bp in length  
\* 158483 158582: gap of 100 bp  
\* 158583 160442: contig of 1860 bp in length  
\* 160443 160542: gap of 100 bp  
\* 160543 161624: contig of 1082 bp in length.  
location/Qualifiers

FEATURES  
source

misc\_feature  
1. 161624  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18q21"  
/clone="RP11-779H19"  
1. 32754  
/note="assembly\_fragment"

misc\_feature  
32855. 48719  
/note="assembly\_fragment clone\_end:77 vector\_slide:left"

misc\_feature  
48820. 64088  
/note="assembly\_fragment"

misc\_feature  
64189. 76256  
/note="assembly\_fragment"

misc\_feature  
76357. 86295  
/note="assembly\_fragment"

misc\_feature  
86396. 94552  
/note="assembly\_fragment"

misc\_feature  
94653. 102595  
/note="assembly\_fragment"

misc\_feature  
102696. 110667  
/note="assembly\_fragment"

misc\_feature  
110768. 116555  
/note="assembly\_fragment"

misc\_feature  
116556. 122099  
/note="assembly\_fragment"

misc\_feature  
122200. 125461  
/note="assembly\_fragment"

misc\_feature  
125562. 131077  
/note="assembly\_fragment"

misc\_feature  
131178. 135924  
/note="assembly\_fragment"

misc\_feature  
136025. 138309  
/note="assembly\_fragment clone\_end:SP6 vector\_slide:left"

misc\_feature  
138410. 140832  
/note="assembly\_fragment"

misc\_feature  
140933. 144613  
/note="assembly\_fragment"

misc\_feature  
144714. 147380  
/note="assembly\_fragment"

misc\_feature  
147481. 150820  
/note="assembly\_fragment"

misc\_feature  
150921. 154114  
/note="assembly\_fragment"

misc\_feature  
154215. 156288  
/note="assembly\_fragment"

misc\_feature  
156389. 158482  
/note="assembly\_fragment"

misc\_feature  
158583. 160442  
/note="assembly\_fragment"

misc\_feature  
160543. 161624  
/note="assembly\_fragment"

BASE COUNT 49551 a 30115 c 30590 g 49168 t 2200 others

ORIGIN

	Query Match	Score 18.4	DB 2	Length 161624
	Best Local Similarity	95.0%	Pred. No. 6.4e+02	
Matches	19	Conservative	0	Mismatches 1, Indels 0, Gaps 0
Oy	2	tgaagaagaagctctaaagt	21	
Db	38706	TGAAAAGCAAGCTTAAAGT	38687	

RESULT	13
AC026923/c	
LOCUS	AC026923
DEFINITION	Homo sapiens clone RP11-725F23, WORKING DRAFT SEQUENCE, 13
ACCESSION	AC026923
VERSION	AC026923.2
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Plrimates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 164652)
AUTHORS	Birten,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens, Clone RP11-725F23
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 164652)
AUTHORS	Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.

**TITLE** Direct Submission  
**JOURNAL** Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Apr 20, 2000 this sequence version replaced gi:7328788. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

```

-----
Center code: MIBR
Center code: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L7522
Center clone name: 725.F.23
-----
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156378 bases at least Q40
Consensus quality: 160493 bases at least Q30
Consensus quality: 162206 bases at least Q20
Insert size: 145000; agarose-fp
Insert size: 16455; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp

```

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	1543:	contly of 1543 bp	in length
*	1544	1643:	gap of 100 bp	
*	1644	3512:	contly of 1869 bp	in length
*		3513:	gap of 100 bp	
*	3613	6325:	contly of 2713 bp	in length
*	6326	6425:	gap of 100 bp	
*	6426	12332:	contly of 5807 bp	in length
*	12233	12332:	gap of 100 bp	
*	12333	17689:	contly of 3537 bp	in length
*	17690	17789:	gap of 100 bp	
*	17790	23842:	contly of 6053 bp	in length
*	23843	23942:	gap of 100 bp	
*	23943	30454:	contly of 6512 bp	in length
*	30455	30554:	gap of 100 bp	
*	30555	39019:	contly of 8465 bp	in length
*	39020	39119:	gap of 100 bp	
*	39120	52290:	contly of 1171 bp	in length
*	52291	52390:	gap of 100 bp	
*	52391	65949:	contly of 15559 bp	in length
*	65950	66049:	gap of 100 bp	
*	66050	84406:	contly of 18357 bp	in length
*	84407	84506:	gap of 100 bp	
*	84507	105983:	contly of 25477 bp	in length
*	105984	110083:	gap of 100 bp	
*	110084	164652:	contly of 54569 bp	in length

BASE COUNT	51861 a	28406 c	28375 g	54786 t	1204 others
misc.feature	1.	1543			
misc.feature	/note="assembly_fragment"	1644.	3512		
misc.feature	/note="assembly_fragment"	3613.	6325		
misc.feature	/note="assembly_fragment"	6426.	12232		
misc.feature	/note="assembly_fragment"	12333.	17689		
misc.feature	/note="assembly_fragment"	17790.	23842		
misc.feature	/note="assembly_fragment"	23943.	30454		
misc.feature	/note="assembly_fragment"	clone_end:SP6			
misc.feature	vector_side:left"	30555.	33019		
misc.feature	/note="assembly_fragment"	39120.	52290		
misc.feature	/note="assembly_fragment"	52331.	65949		
misc.feature	/note="assembly_fragment"	66050.	84406		
misc.feature	/note="assembly_fragment"	clone_end:17			
misc.feature	vector_side:right"	84507.	109983		
misc.feature	/note="assembly_fragment"	110084.	164652		
misc.feature	/note="assembly_fragment"				

Query Match 87.6% Score 18.4; DB 2; Length 164652;  
 Best Local Similarity 95.0%; Pred. No. 6.4e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atcaaaaggaagcctaagaag 20  
 Db 93160 AAAAAAGGAGCTAAG 93141

RESULT 14  
 APO01548/c  
 LOCUS APO01548.2  
 DEFINITION Homo sapiens chromosome 18 clone RP11-83712 map 18q21, WORKING  
 DRAFT SEQUENCE, 17 unordered pieces.  
 ACCESSION APO01548  
 VERSION APO01548.2 GI:8117386  
 KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.  
 SOURCE Homo sapiens DNA, clone:RP11-83712.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 170270)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Homo sapiens 170,270 genomic DNA of 18q21  
 Published Only in Database (2000) In press  
 2 (bases 1 to 170270)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Direct Submission  
 Submitted (24-MAR-2000) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
 Japan (E-mail:hattori@gsc.riken.go.jp,  
 URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,  
 Fax:81-42-778-9924)  
 On May 30, 2000 this sequence version replaced gi:7340850.

COMMENT  
 ----- Genome Center  
 Center: RIKEN Genomic Sciences Center(GSC)  
 Center code: RIKEN  
 Web site: http://hgp.gsc.riken.go.jp/  
 Contact: hattori@gsc.riken.go.jp  
 ----- Project Information  
 Center project name: HumDrafl18  
 Center clone name: RP11-83712  
 ----- Summary Statistics  
 Sequencing vector: PCR products; 100% of reads  
 Chemistry: Dye-terminator ET-amersham; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 162012 bases at least Q40  
 Consensus quality: 165830 bases at least Q30  
 Consensus quality: 167538 bases at least Q20  
 Insert size: 168670; sum-of-contents  
 Quality coverage: 5.46x in Q20 bases; sum-of-contents

NOTE: This is a 'working draft' sequence. It currently consists of  
 17 contigs. The true order of the pieces is not known and their  
 order in this sequence record is arbitrary. Gaps between the  
 contigs are represented as runs N, but the exact sizes of the gaps  
 are unknown. This record will be updated with the finished sequence  
 as soon as it is available and the accession number will be  
 preserved

1 30841 contig of 30841 bp in length  
 30942 49995 contig of 19054 bp in length  
 50096 70667 contig of 20572 bp in length  
 70768 89290 contig of 18523 bp in length  
 89391 105447 contig of 16057 bp in length  
 105548 120373 contig of 14826 bp in length  
 120474 133152 contig of 12679 bp in length  
 133253 143203 contig of 9951 bp in length  
 143304 147902 contig of 4599 bp in length

148003 152506 contig of 4504 bp in length  
 152607 155699 contig of 3093 bp in length  
 155800 158705 contig of 2906 bp in length  
 155806 162039 contig of 3234 bp in length  
 162140 164114 contig of 1975 bp in length  
 164215 165502 contig of 2288 bp in length  
 165603 166602: gap of 100 bp  
 166603 169081: contig of 2479 bp in length  
 169082 169181: gap of 100 bp  
 169182 170270: contig of 1089 bp in length.

Sequence updated (26-May-2000).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 17 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 30841: contig of 30841 bp in length  
 30842 30941: gap of 100 bp  
 30942 49995: contig of 19054 bp in length  
 49996 50095: gap of 100 bp  
 50096 70667: contig of 20572 bp in length  
 70668 70767: gap of 100 bp  
 70768 89290: contig of 18523 bp in length  
 89291 89390: gap of 100 bp  
 89391 105447: contig of 16057 bp in length  
 105448 105547: gap of 100 bp  
 105548 120373: contig of 14826 bp in length  
 120374 120473: gap of 100 bp  
 120474 133152: contig of 12679 bp in length  
 133153 133252: gap of 100 bp  
 133253 143203: contig of 9951 bp in length  
 143204 143303: gap of 100 bp  
 143304 147902: contig of 4599 bp in length  
 147903 148002: gap of 100 bp  
 148003 152506: contig of 4504 bp in length  
 152507 152606: gap of 100 bp  
 152607 155699: contig of 3093 bp in length  
 155700 155799: gap of 100 bp  
 155800 158705: contig of 2906 bp in length  
 158706 158805: gap of 100 bp  
 158806 162039: contig of 3234 bp in length  
 162040 162139: gap of 100 bp  
 162140 164114: contig of 1975 bp in length  
 164115 164214: gap of 100 bp  
 164215 165502: contig of 2288 bp in length  
 165503 166602: gap of 100 bp  
 166603 169081: contig of 2479 bp in length  
 169082 169181: gap of 100 bp  
 169182 170270: contig of 1089 bp in length.

FEATURES  
 source  
 1. 170270  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="18"  
 /map="18q21"  
 /clone="RP11-83712"  
 1. 30841  
 /note="assembly-fragment"  
 30942. 49995  
 /note="assembly-fragment"  
 50096. 70667  
 /note="assembly-fragment"  
 70768. 89290  
 /note="assembly-fragment"  
 89391. 105447  
 /note="assembly-fragment clone\_end:T7 vector\_side:left"  
 105548. 120373  
 /note="assembly-fragment"  
 120474. 133152  
 /note="assembly-fragment clone\_end:SP6 vector\_side:right"  
 133253. 143203  
 /note="assembly-fragment"

```

misc_feature      14304. .147902
                  /note="assembly_fragment"
misc_feature      148003. .152506
                  /note="assembly_fragment"
misc_feature      152607. .155699
                  /note="assembly_fragment"
misc_feature      155800. .158705
                  /note="assembly_fragment"
misc_feature      158806. .162039
                  /note="assembly_fragment"
misc_feature      162140. .164114
                  /note="assembly_fragment"
misc_feature      164215. .166502
                  /note="assembly_fragment"
misc_feature      166603. .169081
                  /note="assembly_fragment"
misc_feature      169182. .170270
                  /note="assembly_fragment"
BASE COUNT      51510 a 32818 c 32593 g 51749 t 1600 others
ORIGIN

```

```

Query Match      87.6%; Score 18.4; DB 2; Length 170270;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 tgaagaagaacttaagt 21
|||||
Db 84709 TCAGAAAGCAAGCTAAGCT 84690

```

```

RESULT 15
AC090405          AC090405 170413 bp DNA HTG 27-MAY-2001
LOCUS            Homo sapiens chromosome 18 clone RP11-779H19 map 18, WORKING DRAFT
DEFINITION       AC090405
SEQUENCE         AC090405.2 GI:14210571
KEYWORDS         HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTROP.
SOURCE           human.
ORGANISM         Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS          Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
TITLE            1 (bases 1 to 170413)
JOURNAL          Homo sapiens chromosome 18, clone RP11-779H19
REFERENCE        Unpublished
AUTHORS          2 (bases 1 to 170413)
                  Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
                  Barina,N., Bastien,P., Boguslavsky,L., Boukhalter,B., Brown,A.,
                  Canarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
                  Collamore,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S.,
                  Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
                  Gardyna,S., Ginde,S., Goette,M., Graham,L., Grand-Pierre,N.,
                  Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
                  Jones,C., Karates,A., Larocque,K., Lamazares,R., Landers,T.,
                  Lencoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
                  McPheters,R., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
                  Murphy,T., Naylor,J., Nguyen,C., Nordu,C., Norman,C.H.,
                  O'Connor,T., O'Donnell,P., O'Neill,J., Oliver,J., Peterson,K.,
                  Phunkhang,P., Plerre,N., Pollara,V., Raymond,C., Retta,R.,
                  Rhinback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
                  Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
                  Sougner,C., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
                  Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
                  Travers,M., Travis,N., Triggillo,W., Vassiliev,H., Viel,R., Vo,A.,
                  Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
                  Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
JOURNAL          Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
COMMENT          Research, 320 Charles Street, Cambridge, MA 02141, USA
                  On May 27, 2001 this sequence version replaced gi:12958041.

```

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://fpi.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WMR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L12714

Center clone name: 779\_H.19

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 165803 bases at least Q40

Consensus quality: 167767 bases at least Q30

Consensus quality: 168509 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 169313; sum-of-contigs

Quality coverage: 8.9 in Q20 bases; agarose-fp

Quality coverage: 8.3 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 27912: contig of 27912 bp in length
* 27913 28012: gap of 100 bp
* 28013 28759: contig of 747 bp in length
* 28760 28859: gap of 100 bp
* 28860 29596: contig of 737 bp in length
* 29597 29696: gap of 100 bp
* 29697 30720: contig of 1024 bp in length
* 30721 30820: gap of 100 bp
* 30821 31522: contig of 702 bp in length
* 31523 31622: gap of 100 bp
* 31623 32368: contig of 746 bp in length
* 32369 32468: gap of 100 bp
* 32469 33330: contig of 862 bp in length
* 33331 33430: gap of 100 bp
* 33431 34637: contig of 1207 bp in length
* 34638 34737: gap of 100 bp
* 34738 38948: contig of 4211 bp in length
* 38949 39048: gap of 100 bp
* 39049 43570: contig of 4522 bp in length
* 43571 43670: gap of 100 bp
* 43671 162170: contig of 118500 bp in length
* 162171 162270: gap of 100 bp
* 162271 170413: contig of 8143 bp in length.

```

----- Location/Qualifiers

1. 170413

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-779H19"

/clone\_11b="RPC1-11 Human Male BAC"

1. 27912

/note="assembly\_fragment"

clone\_end:sp6

vector\_side:left"

28013. .28759

/note="assembly\_fragment"

28860. .29596

/note="assembly\_fragment"

29697. .30720

/note="assembly\_fragment"

30821. .31522

misc\_feature

misc\_feature

misc\_feature

```

misc_feature      /note="assembly_fragment"
31623..32368
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
32469..33330
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
33431..34637
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
34738..38948
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
39049..43570
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
43671..162170
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
162271..170413
/note="assembly_fragment"
clone_end:T7
vector_side:right"

```

BASE COUNT 50853 a 32105 c 32644 g 53709 t 1102 others  
ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 170413;  
Best Local Similarity 95.0%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaagaaggaagcctaaagt 21  
|||||  
Db 164562 TGAAGAAGCAAGCTAAAAAGT 164581

Search completed: March 20, 2002, 01:59:19  
Job time: 8555 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 23:03:34 ; Search time 583.26 Seconds  
(without alignments)  
44.535 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719  
Sequence: 1 MKRKASILEFLSTVLFQAQ.....FQNFVEKIESKPEESSPKN 341

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	99.0	344	2	C70183 flagellar filament
2	361.5	21.0	350	2	D71348 flagellar filament
3	331.5	19.3	320	1	A43824 periplasmic flagel
4	202.5	11.8	301	2	A32814 flagellar filament
5	121.5	7.1	242	2	H71297 probable outer mem
6	112	6.5	4152	2	T31102 filamentous hemaggl
7	109	6.3	444	1	I40492 gntk-type transcri
8	108	6.3	580	2	F64385 hypothetical prote
9	106.5	6.2	695	2	G64327 H+-transporting AT
10	106	6.2	1181	2	C82500 icmF-related prote
11	105.5	6.1	705	2	A41322 N-acetylmuramoyl-L
12	104.5	6.1	1119	2	B70126 surface-located me
13	104.5	6.1	4919	2	T31105 hypothetical prote
14	104	6.1	499	2	D67335 xylofucose kinase xy
15	103	6.0	650	2	S14949 inter-alpha-trypsin
16	102.5	6.0	914	2	JC5574 probable flagellar
17	102	5.9	241	2	A71298 mosquitoicidal prot
18	101.5	5.9	613	2	JC6033 hypothetical prote
19	101.5	5.9	622	2	A64494 site-specific DNA-
20	101.5	5.9	627	2	B64710 dnak-type molecule
21	101.5	5.9	676	2	A45515 major meroclit su
22	101	5.9	1060	2	S06286 hypothetical prote
23	101	5.9	2166	2	G70163 hypothetical prote
24	100.5	5.8	318	2	E71240 unknown protein F9
25	100	5.8	651	2	A96781 glycine-tRNA L1ga
26	100	5.8	701	2	H71874 hypothetical prote
27	99.5	5.8	401	2	A69368 inter-alpha-trypsi
28	99.5	5.8	907	2	S54353 hypothetical prote
29	99	5.8	1408	2	T45039

30	98.5	5.7	419	2	A56153 46K surface antige
31	98.5	5.7	644	2	S14950 dnak-type molecule
32	98.5	5.7	651	2	T45517 heat shock protein
33	98	5.7	635	2	S61175 probable membrane
34	98	5.7	707	2	S57157 hypothetical prote
35	98	5.7	913	2	D82885 multiple banded an
36	97.5	5.7	425	2	C64567 fucosyltransferase
37	97.5	5.7	721	2	A33319 microtubule-associ
38	97	5.6	564	2	T00135 hypothetical prote
39	97	5.6	1136	2	T40355 hypothetical prote
40	97	5.6	3848	2	T17414 TIPC protein - sli
41	96	5.6	1875	2	S38173 myosin-like protei
42	95.5	5.6	634	2	T01408 inclusion protein
43	95.5	5.6	647	2	T45522 heat shock protein
44	95.5	5.6	651	2	S46302 dnak-type molecule
45	95.5	5.6	651	2	T48271

## ALIGNMENTS

RESULT 1  
C70183  
flagellar filament outer layer protein (flaA) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 02-Sep-2000  
C:Accession: C70183  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
son, D.; Peterson, J.; Kierlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu  
Boman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; M0ID:98065943  
A:Accession: C70183  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-344 <L>  
A:Cross-references: GB:AE001168; GB:AE000783; NID:g2688598; PID:AA67025.1; PID:g268  
A:Experimental source: strain B31  
C:Superfamily: spirochete flagellar sheath protein

Query Match 99.0%; Score 1701; DB 2; Length 344;  
Best Local Similarity 98.8%; Pred. No. 5.8e-116;  
Matches 337; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY	1	MKRRKASILEFLSTVLFQAQETDGLAEGSKRAEPGELVDFPAELARPPSTRDLTNYVD	60
DB	4	MKRRKASILEFLSTVLFQAQETDGLAEGSKRAEPGELVDFPAELARPPSTRDLTNYVD	63
QY	61	YVSGASGIKPEDMVVDLGINNMSVLLTPSARLQAVKNSVAVPAVAKSESKRYAGDTI	120
DB	64	YVSGASGIKPEDMVVDLGINNMSVLLTPSARLQAVKNSVAVPAVAKSESKRYAGDTI	123
QY	121	LGVRVLPFSYSQSSAMIMPFKIPFISGEGNOFLGGLDNIKTMEIVSVSLGEYI	180
DB	124	LGVRVLPFSYSQSSAMIMPFKIPFISGEGNOFLGGLDNIKTMEIVSVSLGEYI	183
QY	181	DLVLVLFEDMNGMYAVSMGTLKFGNADLIMSNNYIPNISRIKIDVNNYPLASSKM	240
DB	184	DLVLVLFEDMNGMYAVSMGTLKFGNADLIMSNNYIPNISRIKIDVNNYPLASSKM	243
QY	241	FKAFRVSKSHSKVKNEIFVYKDLRVLYDKLSVSDSDISESVFKYETSGTESLRKLK	300
DB	244	FKAFRVSKSHSKVKNEIFVYKDLRVLYDKLSVSDSDISESVFKYETSGTESLRKLK	303
QY	301	AHETFKRVLLRKIKISAESSFQNFVEKIESKPEESSPKN 341	
DB	304	AHETFKRVLLRKIKISAESSFQNFVEKIESKPEESSPKN 344	

RESULT 2

071348  
Flagellar filament outer layer protein flaA-1 precursor - syphilis spirochete  
NAlternate names: 37k endoflagellar sheath protein FlaA; minor endoflagellar protein; F  
CSpecies: Treponema pallidum subsp. pallidum (syphilis spirochete)  
CDate: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 17-Mar-2000  
CAccession: D71348; A40160; A7187; B27559; A32351  
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
rson, J.; Khaliq, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo  
ney, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770  
A:Accession: D71348  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-350 <COL>  
A:Cross-references: GB:AE001206; GB:AE000520; NID:g3322514; PIDN:AA65235.1; PID:g332251  
A:Experimental source: strain Nichols  
R:Isacs, R.D.; Radolf, J.D.  
I:fect. Immun. 58, 2025-2034, 1990  
A>Title: Expression in Escherichia coli of the 37-kilodalton endoflagellar sheath protei  
A:Reference number: A40160; MUID:90307197  
A:Accession: A40160  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-29 <ISA1>  
A:Cross-references: GB:M63142; GB:M27915; NID:g155058  
R:Isacs, R.D.; Hanke, J.H.; Guzman-Verduzco, L.M.; Newport, G.; Agabian, N.; Norgard, M  
I:fect. Immun. 57, 3403-3411, 1989  
A>Title: Molecular cloning and DNA sequence analysis of the 37-kilodalton endoflagellar  
A:Reference number: A37187; MUID:90035409  
A:Accession: A37187  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 21-350 <ISA2>  
A:Cross-references: GB:M26525; NID:g155056  
R:Blanco, D.R.; Champion, C.I.; Miller, J.N.; Lovett, M.A.  
I:fect. Immun. 56, 168-175, 1988  
A>Title: Antigenic and structural characterization of Treponema pallidum (Nichols strain  
A:Reference number: A91753; MUID:88085470  
A:Accession: B27559  
A:Molecule type: protein  
A:Residues: 21-28, 'X', 30-36, 'G', 38-40 <BLA>  
R:Norris, S.J.; Charon, N.W.; Cook, R.G.; Fuentes, M.D.; Lamberger, R.J.  
J. Bacteriol. 170, 4072-4087, 1988  
A>Title: Antigenic relatedness and N-terminal sequence homology define two classes of p  
A:Reference number: A91882; MUID:88314903  
A:Accession: A32351  
A:Molecule type: protein  
A:Residues: 21-48, 'R', 50-51 <NOR>  
C:Genetics:  
A:Gene: TP0249  
C:Superfamily: spirochete flagellar sheath protein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-350/Product: flagellar filament outer layer protein flaA-1 #status experimental <MA

```

QY      163  IKTKEIVSVSYSLGIEIDLEVLFDNGMGMRYAAVSMGLTKFKNAADILMSNPNIPINSS 222
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      178  VGYLKSIAVNTYGMNYPHGLLYVMARDQGEVYHRIFPMGLIFLDSKMLVWNNPSTISDVS 237
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      223  RIITKDVPENYPLASSKMKRFKFRVSKSSHSSKVKNEFIYKDLRVLYDKLSVSDSIDSE 282
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      238  REVR-LYPIYPASPSTHYVEEGPMVTRDAAHAGDVGVEYFKDKVITIKYDKAVLSTVRFADE 296
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      283  SVFKYVEFTSGTESLRLKLAHETFRKVLKREKISLAEBSFQNFPEKITESEK 333
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      297  DLW-----GIQARREAE-----RRKVEYARFGQOQVLRKTYIQEK 330
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT      3
A43824
periplasmic flagellar sheath protein precursor - Treponema hyodysenteriae
C.Species: Treponema hyodysenteriae
C.Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C.Accession: A43824
R.Koopman, M.B.; de Laeuw, O.S.; van der Zielst, B.M.; Kuusters, J.G.
Infect. Immun. 60, 2920-2925, 1992
A.Title: Cloning and DNA sequence analysis of a Serpulina (Treponema) hyodysenteriae
A.Reference number: A43824; MUID:92307926
A.Accession: A43824
A.Molecule type: DNA
A.Residues: 1*320 <KOO>
A.Cross-references: EMBL:X63006; NID:9433523; PIDN:CAA44735.1; PID:9433524
C.Comment: The flagella of this spirochete are reported to be enclosed completely with
C.Genetics:
A.Gene: flaA
C.Superfamily: spirochete flagellar sheath protein
C.Keywords: flagellum; periplasmic space
E1-19/Domains: signal sequence #status predicted <SIG>
E1-20-320/Product: periplasmic flagellar sheath protein #status experimental <MAT>

```

```

Query Match      19.3%; Score 331.5; DB 1: Length 320;
Best Local Similarity 28.8%; Pied. No. 9,4e-17;
Matches 90; Conservative 63; Mismatches 129; Indels 31; Gaps

QY 9 LEFLLSTVLVLFQNEFDGLAESGRKRAPEGLVDLDAEFLARDPSSRLDLINNYDYYVSGASG 68
   ||:::|  ||  ::||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||
DB 4 LEVVTLSIFFLAASKYGLTNSI-----LIDRA-----LTGMNDNLQAGSG 43

QY 69 IYKPEDMYVDGIGNNWSVLLTPSARLQAYVKNVSVAPAYKSESKRYAGDTLLGVRVLP 128
   ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||
DB 44 TNEVYPVAENLYNDNMVYWLINESARLTENRRSNRYTNVDVSKNNQAMEGAKVLGYRVHP 103

QY 129 SYSQSS-AMIMPPKIPFYSGESGNQL-CKGLIDNIKMKKEIKSVYSLGEIDLEVL 186
   ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||
DB 104 LAAMNSYVLVRYVELMYGGADQTKTECKGIHNVGEEKISISSWYGRNLLISYFNL 163

QY 187 EDMNMEVAAYSMTGLFKFGNADLTWSNPNIPISSRIITKDDVPNYPVPLASSMKRPKARV 246
   ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||
DB 164 QNEBELKSYMGVYINNGRQVMEKRELPLPVRRVRLVRE-PLXPRIPIPVKLDISGF 222

QY 247 SKSHSSKYKNEIFYKDLRLVLDKLSYISDSDISESVKRYEYTSQTESLRKLKAHETFK 306
   ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||
DB 223 YRTKDTKCGDFTYVKDVTLEYDVVVVDFFEDIDDEATWQLK---TENDRK-QAIES-- 276

QY 307 RYVLKLEKISIAE 319
   ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||  ::||  ||
DB 277 --ARIREQAELRD 287

RESULT 4
A52814
flagellar filament surface antigen - Spirochaeta aurantia
C:Species: Spirochaeta aurantia
C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 07-Feb-1997
C:Accession: A52814
J:Bransham, B.; Greenberg, E.P.
J: Bacteriol. 171, 1692-1697, 1989

```



A:Title: Cloning and sequence analysis of flaA, a gene encoding a Spirochaeta aurantia fl  
A:Reference number: A32814; MUID:89155480  
A:Accession: A32814  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-301 <BNA>  
C:Keywords: surface antigen

Query Match 11.8%; Score 202.5; DB 2; Length 301;  
Best Local Similarity 31.0%; Pred. No. 1.9e-07;  
Matches 57; Conservative 32; Mismatches 68; Indels 27; Gaps 7;

QY 1 MKRKASILEFLFPAQETDGLAEGSKRAEPGELVDFAEIARDPSSRLDITNVYD 60  
DB 1 MKR-----FFAIGAALEFGNSGAFAEQA-----TLIDFSKIVGE-GNTGLHAPPTID 47  
QY 61 YVSGASGIKPED---MVVDLGINMNSVLLTPSARLQAVKNSVPAVAVKSEKRYAG 117  
DB 48 YSRQAGSA-YSAEDKAMKISLAIPSEIELASSQTVENQTLSTVTAAPVKKDAAHYGG 106  
QY 118 DTLGVRVLFPSYS-OSSAMIMPPFKIPY-----SGESGNQFLGKGLDNIKTMK 167  
DB 107 ETVMGVRHPPSPGINSFAVIKPPFTIPAYATLGDATQAQNAVAGGDFGFGVLKMGVVK 166  
QY 168 EIKV 171  
DB 167 SIQI 170

RESULT 5  
H71297  
probable outer membrane protein - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: H71297

R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwalt

ron, J.; Khakh, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 261, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: H71297

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-242 <COL>

A:Cross-references: GB:AE001240; GB:AE000520; NID:g3322952; PIDN:AAC65636.1; PID:g332296

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0663

Query Match 7.1%; Score 121.5; DB 2; Length 242;  
Best Local Similarity 22.0%; Pred. No. 0.1;  
Matches 54; Conservative 31; Mismatches 105; Indels 55; Gaps 8;

QY 48 PSSTRDLITNVYV---YSGAGIYKPEDMVVDLGINM---SVLLTPSARLQAVKN 100  
DB 21 PLAORSKVYQAVFIIDDFDGS-----EDQGL-AMRAAGSKFTKGFPIKRYEG 70  
QY 101 SVAPAVVAVKSEKRYAGDTLIGVRVLFPSYSOSSAMIMP-----PKIIPFSGSSGNFL 155  
DB 71 MPQAVRAGSWGCKKKEARIGVECKENRGNMMLDILPIKGSVDYELPLRGVSG--- 126  
QY 156 GKGLDNIKTMKEIKVSVLSYEIDLEVLFEIDNGMEVAVYSMTLKFKNADLIWSPN 215  
DB 127 -----FDVWVWAGVQYSLEALVROCTGRVHLLIGNLDFQCKKNLSVGPPT 173  
QY 216 YIPNISRIIKDVPNPLASSKMRFAFVSKSHSKVKNFTFYVKDLRYLKLVSIV 275  
DB 174 HIPTI-----RYLGSQHLSEFVGFRIRTXPSERVDVY-----VFEQFALA 217  
QY 276 DSDID 280

DB 218 NMHID 222

RESULT 6  
T31102  
filamentous hemagglutinin 1 - Haemophilus ducreyi

C:Species: Haemophilus ducreyi

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T31102

R:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.

J. Bacteriol. 180, 6013-6022, 1998

A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.

A:Reference number: Z20984; MUID:99030326

A:Accession: T31102

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4152 <WAR>

A:Cross-references: EMBL:AF057695; NID:g3292017; PID:g3292018; PIDN:AAC79757.1

C:Genetics:

A:Gene: 18PA1

Query Match 6.5%; Score 112; DB 2; Length 4152;  
Best Local Similarity 22.8%; Pred. No. 28;  
Matches 84; Conservative 44; Mismatches 133; Indels 108; Gaps 15;

QY 4 KAKSILEFLFPAQETDGLAEGSKRAEPGELVDFAEIARDPSSRLDITNVYD 63  
DB 12 KVKNCIVPAVENIKSASGSSGSSNSKIAEDQ-----EEDPSIACSLSS----- 55  
QY 64 SGAGIYKPEDMVVDLGINMNSVLLTPSARLQAVKNSVPAVAVKSEKRYAG 97  
DB 56 -----PLSSSTHLGHNHSPKVEKGLSVLLSLMPATPLAQAQNAEALNGKV 107

QY 98 VKNVPAVAVKSEKRYAGDTLIGVRVLFPSYSOSSAMIMPPKIP-----FYSQ-ESG 151  
DB 108 VDSOHSSTRIDEQNTDNSKGIYVELANEDVGSDNRKKEFNIPSAFNNSTRIST 167

QY 152 NQFLGKGLDNIKTMKEIKVSVLSYEIDLEVLFEIDNGMEVAVYSMTLKFKN- 210  
DB 168 SOLVGK-LHANIQIQEKVK-----LILNVTGHNESNIQALFVAGKAKADI 213

QY 211 WSNPNYIPNISRIIKDVPNPLASSKMRFAFVSKSHSKVKNFTFYVKDLRY 266  
DB 214 IVNNGTTLNGVKRTIND-----RRVSTSDIIPR--ENGILLVRNCKV 256

QY 267 LYDRLYS-----IDSDISESVKYEYETGTESLRKLAHEFP-----KRVTLKR 312  
DB 257 TIDKGVATNGLSHFEVAVARNIDQKRTTAKTENQSVN--PANITFAAGSLNINLKR 314

QY 313 EKISIAEGS 321  
DB 315 EATPISSCT 323

RESULT 7  
I40492  
gnt-type transcription regulator homolog ycdX - Bacillus subtilis  
N:Alternate names: hypotheoretical protein 8 (grfA operon)  
C:Species: Bacillus subtilis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: I40492; F69766; S34990  
R:Cosmina, P.; Rodrigue, F.; de Ferro, F.; Grandi, G.; Perego, M.; Venema, G.; van S  
Mol. Microbiol. 8, 821-831, 1993  
A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthe  
A:Reference number: I40485; MUID:93360813  
A:Accession: I40492  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-444 <RES>  
A:Cross-references: EMBL:X70356; NID:g396480; PIDN:CAA49821.1; PID:g396486  
A:Experimental source: ATCC 21332

```

Query Match          6.3%; Score 109; DB 1; Length 444;
Best Local Similarity 21.6%; Pred. No. 2;
Matches 75; Conservative 42; Mismatches 126; Indels 104; Gaps 15;

QY      29 SKRAREGELVLDFAEFLARDPSSRLDTLNTNYD-----YYVSGASGI--- 69
       || :|| :||| | | | | | | | | | | | | | | | | | | | | | | |
Db      73 SKSQGPGR--IDFATNSADP-----DVFPEIDFGHCINKAIDFYKNDLFITYGTPLGLSL 125
OY      70 -----VKPEDMVVDLCINMKSVYL-----TPSARKLDAYKN 100
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      126 IRYLRLLATQQVFADERHIFITTSVGQAALSLCAMPENCKEIALIDPGYHLWVEGLE 185
OY      101 SYVAAPAVKSESKRRACPTILICRVLPSPYSOSSAMIMPPKRIPIFYSGBNQIIGKGLI 160
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      186 TLGIPAIQIKRKEE--GDIADIAPERLFOETSIKIIFYTMRFRINP-----LGCSLS 233
OY      161 DNIKTKMEIKSVSYLYGEIDLEVLPEDMNGMEYVSMGTLEFKGMADLLMSNPYIIPNI 220
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      234 ERDKO-----ELVRLAAAYDYVLYVEDDY-----LGDLEENKKAD----PLYAYDL 274
OY      221 SSRITIKDVPNPPLASSKMRFKAFFVSXSHSKVKNFIIFYKVDLRLVLDKLSYSTDSIID 280
       || :|| :||| | | | | | | | | | | | | | | | | | | | | | | |
Db      275 SSHVI-----YLKFSFKMMPEGLRVGAAYVLEALTDTFFYA-----YKKLN-DIDCSI 321
OY      281 SESVRYKYETSGETESLRKLKAHEFER-RVCLKPREKI-----STAGESSF 322
       || :|| :|| | | | | | | | | | | | | | | | | | | | | | | |
Db      322 SQAALETIYAKSCMYGRHKKEKITDSYKERSLRHQAIRTHROLGSGRF 368

RESULT      8
FE64385
hypothetical protein MJ0686 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Oct-1999
C:Accession: F64385
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
J.; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Glodek, A.;
Iron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: FE64385
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-580 <BU>
```

```
Query Match      6.3%; Score 108; DB 2; Length 580;  
Best Local Similarity 21.7%; Pred. No. 3.4;  
Matches 56; Conservative 38; Mismatches 66; Indels 98; Gaps 11;
```

OY	112	SRRVAGDTLLGVRVLPPSTSSSAMIMPEKIPETYSGGSGNOFLCKGLINDIKTKKELV	171
Dz	376	SERYSGST-----ESRKMKDFEVSQWIMNNKI--PEVENIKI--EIoy	417
Oy	172	SVYESGEIDLEVLFE DMNMGEMAYSMGTLFKFGADLTWSNP--YIPNISRIIDD	228
Dz	418	GAVSL-----KKVASGSSSKKGVFNIIFPKKPMPDFFKP-----DN	453
Oy	229	VPNYPLASSKMFKAF-----RVSKSHSRYKNPIFYK	262
Dz	454	IAYVYLEDHHPKFGLRNKGISNEYIDSVLNKTPILDJTNKKISKSPSK----	YVK 507
Oy	263	LDRVLVDKLVSIDSIDI-----SEGVFVEYTESGESLPRLKAHETFPRVLKL	311
Dz	508	KMIIELOKKNGISEDVAVNKYRILELGHFIINEMFILANT--DDLSLKOEIDENFRFIEL	566
Oy	312	REKISIABGSFOYVEKI	329
Dz	567	REKL-----ILEKI	575

RESULT 9  
 G64327  
 H<sup>+</sup>-transporting ATP synthase (EC 3.6.1.34) chain I - *Methanococcus jannaschii*  
 C:Species: *Methanococcus jannaschii*  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: G64327  
 R:Built, C.J.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 R.; Reich, C.L.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese  
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
 A:Reference number: A64300; MUID:96337999  
 A:Accession: G64327  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-695 <BU>  
 A:Cross-references: GB:U67478; GB:L77117; NID:g1590958; PIDN:AAB98208.1; PID:g159096060  
 C:Genetics:  
 A:Map position: REV214151-212064  
 A:Start codon: GTG  
 C:Keywords: hydrolase

	Query Match	6.2%	Score 106.5;	DB 2;	Length 695;	
	Best Local Similarity	23.5%;	Pred. No. 5.7;			
	Matches	84;	Conservative	41;	Mismatches 108;	Indels 125; Gaps 19;
Qy	68	GIVKPEDNVVDIGINWVSLLTPPSARLQAYKNSVYAPAVYKSESKRYAGDTILGRLV-	126			
		:   :     :   :   :   :				
Dd	32	GIVECLDLSERKELEDLEMKLTLLSPSSAD-YVRN--VTSLIMK-----AG-----RILD	76			
Qy	127	-PPSYQSASAM-----MPPKRIPFYSGSGNOFLKGCL-----ID	161			
		:   :   :   :   :   :				
Dd	77	MESSYSQKETSTKIDLNPPEVKKKRVSNFSNYOEVIDYAEKVLNEISKVEDGPAERSELSD	136			
Qy	162	NIKT-MKEIKSV-YSLGEYIDLEVLFEDMMNGMEYAY-----SMGTLLFKGWADLI	210			
		:   :   :   :   :   :   :   :   :				
Dd	137	NKKSALLQLQKEIDISYLKGLEPDKYL-----GSGEYVFIGAQSVPREKGEELK---AELD	188			
Qy	211	WSNPNTIPNIS-SRIIKDDVPNYPLASSKMRFKAPRVSKSHSXYKANITFYVKDLRLVLD	269			
		:   :   :   :   :   :				
Dd	189	KVADVDYGIFGSSSEFEKKKIKNPIVEVTLKEELENNV---LSELRKEFEREYDI-----	239			

[illegible]

```

Query Match 5: 62%; Score 106; DB 2; Length 1181;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 77; Conservative 54; Mismatches 118; Indels 82; Gaps 17;

QY 28 GSKRAEPGELVLDAFEL-----ADPSSTRLDTNYYVYVS---GASG 68
      |||  |::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 866 GSKPAYMTEVLAVDKLSTKLSIQDAPDVGMALDATTAKRYKLVN-ADPITTLKRISG 924
QY 69 IYKPED-MVVDLGINNMVSVLLTPSARLQAVYKNSVY--APAVKSESKRYAG----- 117
      |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |
DB 925 LKRPDLSMAKKLADSESYVY-----KQEAIKHLEVRKTEDEVYKTFQSGLAGYPPNPAS 978
QY 118 --DTILGVRLLF--PS-----YSQSSAMIMPPEKIPYSGSGNQGLKGLDINIKTMK 167
      |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |
DB 979 NKDVALADEAFAPFAPNCTLDNFYVNOQLKMFIDE-NISVAPSDSAQSIIRKEVLDQIKQAO 1037
QY 168 EIKVSYVSLCYEIDLEVLF-----DMNGEHYVSMGTLKKGADLIWSNP 214
      |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |
DB 1038 KIREAFNKRGIIDVSEFSVPLSLNNKRRSVLVNVDGQFLAYSHGP--RENVETLWPNPT 1094
QY 215 -----NYIP-----NISRIIKDVPNYPYPLASSKMKRFKARYSKSHKKNFIIFYK 262
      |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |
DB 1095 LRDSAVSKYTLTITQTNMSRSIDQ---IQCPWAFPRLLDQGDVYASQTS--VDEKFIYD 1149
QY 263 DLRLVLDKLSVSDSIDSESVFKRYETSGT 293
      |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |::  |
DB 1150 GGEMIV-RINAADANPETERLERLFRKFKLSKT 1179

RESULT 11
A41332
N-acetylglutamate-L-alanine amidase modifier precursor lytB - Bacillus subtilis
N:Alternate names::major autolysin lytC modifier lytB; probable cell wall-binding protein
C:Species: Bacillus subtilis
C:Date: 05-Jun-1992 #sequence_revision 10-Feb-1995 #text_change 15-Oct-1999
C:Accession: A41332; C47679; D69654
R:Kuroda, A.; Sekiguchi, J.
J. Bacteriol. 173, 7304-7312, 1991
A:Title: Molecular cloning and sequencing of a major Bacillus subtilis autolysin gene.
A:Reference number: A41332; M01D:92041629

```

A:Accession: A41322  
A:Molecule type: DNA  
A:Residues: 473-705 <KUP>  
A:Cross-references: GB:M61747  
J.Lazarevic, V.: Margot, P.; Soldo, B.: Karamata, D.  
J. Gen. Microbiol. 138, 1945-1961, 1992  
A:Title: Sequencing and analysis of the *Bacillus subtilis* lyrBAC divergon: a regulat  
A:Reference number: A47679; MUID:93018998  
A:Accession: C47679  
A:Molecule type: DNA  
A:Residues: 1-705 <LAZ>  
A:Cross-references: GB:M87645; NID:g143155; PIDN:AAA22580.1; PID:g143158  
A:Experimental source: strain 168  
A:Note: sequence extracted from NCBI backbone (NCBIN:116645, NCBI:116648)  
R.Kunst, F.: Ogatawa, N.; Moszer, I.: Albertini, A.M.; Alloni, G.: Azevedo, V.: Ber  
C.: Bron, S.: Brouillet, S.: Brusch, C.V.: Caldwell, B.: Capuano, V.: Carter, N.M.:  
A.: Ehrlich, S.D.: Emerson, P.T.: Enliat, K.D.: Errington, J.: Fabel, C.: Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.: Fritz, C.: Fujita, M.: Fujita, S.: Fuma, S.: Galizzi, A.: Gal  
leoch, T.: Harwood, C.R.: Henaut, A.: Hilbert, H.: Holtsappel, S.: Hosono, S.: Hullo, M  
koetter, P.: Konigsstein, G.: Kropp, H.: Kurita, K.: Lapidos, A.: Lardino, S.: Mau  
A:Authors: Lauber, J.: Lazarevic, V.: Lee, S.M.: Levine, A.: Liu, H.: Masuda, S.: Mau  
Y., M.: Ogawa, K.: Ogilawa, A.: Oudega, B.: Park, S.H.: Parro, V.: Pohl, T.M.: Potteece  
Rieger, M.: Rivolta, C.: Rocha, E.: Roche, B.: Rose, M.: Sadale, Y.: Sato, T.: Scall  
A:Authors: Schleif, S.: Schroeter, R.: Scoffone, F.: Sekiguchi, J.: Sekowska, A.: Se  
akeuchi, M.: Tamakoshi, A.: Tanaka, T.: Terpstra, P.: Tognoni, A.: Tosato, V.: Uchiya  
T.: Miners, P.: Mipet, A.: Yamamoto, H.: Yamane, K.: Yasumoto, K.: Yata, K.: Yoshida  
A:Authors: Yoshikawa, H.F.: Zumbstein, E.: Yoshikawa, H.: Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A:Reference number: A69380; MUID:96044033  
A:Accession: D69654  
A:Status: preliminary: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-705 <KUN>  
A:Cross-references: GB:29122; GB:AL009126; NID:g2636029; PIDN:CAB15580.1; PID:el1844  
A:Experimental source: strain 168  
C:Comment: This protein enhances the activity of the N-acetylmutamoyl-L-alanine amidan  
C:Genetics:  
A:Gene: lyrB  
A:Start codon: TTG  
C:Keywords: cell wall  
F1-25/Domian: signal sequence #status predicted <IG>  
F1-26-705/Product: N-acetylmuramoyl-L-alanine amidase modifier #status predicted <MAT>

[illegible]





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 22:13:23 ; Search time 632 Seconds  
(without alignments)  
39.967 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719  
Sequence: 1 MKRRAKSLFLFLSTLVLFNQ.....FQNFVEKIESEKPESSPKN 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_1101:\*

- 1: /SIDS2/gcgdata/geneseq/AA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/AA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/AA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/AA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/AA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/AA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/AA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/AA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/AA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/AA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/AA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/AA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/AA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/AA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/AA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/AA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/AA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/AA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/AA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/AA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/AA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1719	100.0	341	20	AAV28460
2	1645	95.7	335	20	AAV20146
3	1551	90.2	315	20	AAV20147
4	324.5	18.9	320	14	AAV20147
5	125.5	7.3	242	18	AAV22133
6	110	6.4	1657	20	AAV22133
7	110	6.4	1805	20	AAV22133
8	109	6.3	444	14	AAV22133
9	107	6.2	947	19	AAV22133
10	107	6.2	947	20	AAV22133
11	104.5	6.1	1087	20	AAV22133

12	104.5	6.1	1119	20	AAV19934
13	102.5	5.9	391	22	AAV82672
14	101.5	5.9	450	17	AAV82672
15	98.5	5.7	419	17	AAV82672
16	98	5.7	449	22	AAV82672
17	98	5.7	449	22	AAV82672
18	97.5	5.7	530	21	AAV82672
19	97.5	5.7	530	21	AAV82672
20	97.5	5.7	530	21	AAV82672
21	97.5	5.7	530	21	AAV82672
22	97.5	5.7	530	21	AAV82672
23	97.5	5.7	713	13	AAV82672
24	96.5	5.6	419	13	AAV82672
25	95.5	5.6	931	13	AAV82672
26	94	5.5	415	21	AAV82672
27	94	5.5	536	13	AAV82672
28	94	5.5	1382	20	AAV82672
29	93.5	5.4	406	18	AAV82672
30	93.5	5.4	406	18	AAV82672
31	93.5	5.4	406	18	AAV82672
32	93.5	5.4	790	19	AAV82672
33	93.5	5.4	790	20	AAV82672
34	93	5.4	282	10	AAV82672
35	93	5.4	524	20	AAV82672
36	93	5.4	553	20	AAV82672
37	93	5.4	613	21	AAV82672
38	92.5	5.4	789	19	AAV82672
39	92.5	5.4	789	19	AAV82672
40	92.5	5.4	789	20	AAV82672
41	92.5	5.4	789	20	AAV82672
42	92.5	5.4	823	19	AAV82672
43	92.5	5.4	1041	21	AAV82672
44	92	5.4	502	12	AAV82672
45	92	5.4	517	22	AAV82672

#### ALIGNMENTS

RESULT 1	AAV28460	AAV28460 standard; Protein; 341 AA.
ID	AAV28460	
AC	AAV28460	
DT	19-OCT-1999	(first entry)
DE	P37/Flia protein antigen.	
KW	P37; Flia; Lyme disease; antigen; assay; spirochete.	
OS	Synthetic.	
OS	Borrelia burgdorferi.	
FT	Key	Location/Qualifiers
FT	Peptide	1..22
FT	Protein	/label= signal_peptide
FT		23..341
FT		/label= mature_protein
PN	W09935272-A1.	
PD	15-JUL-1999.	
PF	06-JAN-1999;	99WO-US00196.
PR	08-JAN-1998;	98US-0004395.
PA	(INMR) BIOMERIEUX INC.	
PI	Gilmore RD, Johnson BJB;	
DR	WPI; 1999-430399/36.	

B. burgdorferi ant  
S. epidermidis ope  
H. pylori GHP0 109  
Mycoplasma 46-48 k  
Corynebacterium 91  
C. glutamicum prote  
Ehrlichia sp. exte  
Ehrlichia antigen  
Ehrlichia sp. W20.  
Ehrlichia sp. exte  
Canine zona pelluc  
Sequence of surfac  
Saccharomyces cere  
Arabidopsis thalia  
Antigen mc-5c. El  
A. pleuropneumonia  
Sequence of pre-pr  
Human renin. Homo  
Alpha-amylase gene  
Bacillus thuringie  
Bacillus thuringie  
Sequence encoding  
B. burgdorferi ant  
B. burgdorferi ant  
Human colon cancer  
Bacillus thuringie  
Bacillus thuringie  
Bacillus thuringie  
H. pylori GHP0 122  
Plasmodium falcipa  
HincII restriction  
S. cerevisiae apopt

DR	N-PSDB: AAX89520.	
XX	Recombinant Flaa/P37 protein antigen, useful in improved detection	
PT	of Borrelia burgdorferi	
XX		
PS	Claim 2; Page 26-28; 34pp; English.	
XX		
CC	The sequence is the recombinant protein antigen p37/Flaa. Flaa is an	
CC	outer sheath protein of the periplasmic flagella of Borrelia burgdorferi.	
CC	The p37 protein product can be used as an assay for the detection of	
CC	Borrelia burgdorferi infection, the causative agent of Lyme disease. Flaa	
CC	is a prominent antigen in the early humoral immune response to Borrelia	
CC	burgdorferi infection, and significantly suitable for use in improved	
CC	serologic tests for exposure to Lyme disease spirochetes. The detection	
CC	of Flaa can augment the set of recombinant molecules that are recognized	
CC	early in the course of disease and contribute to improved sensitivity of	
CC	early testing for Lyme disease.	
XX		
SO	Sequence 341 AA:	
Query Match	100.0%: Score 1719; DB 20; Length 341;	
Best Local Similarity	100.0%: Pred. No. 6.2e-162;	
Matches 341; Conservative	0; Mismatches 0; Indels 0; Gaps 0	
QY	1 MKRKASLIEFLLSTVLVPEQETDGLAEGRRAEPGEVLDPFALARPSSSTRLLDNTNYD 60	
DB	1 MKRKASLIEFLLSTVLVPEQETDGLAEGRRAEPGEVLDPFALARPSSSTRLLDNTNYD 60	
QY	61 YVYSGASGIVKPEDMVDGGINNWSVLPLPSARLQAVKNSVAPAVYKSESKRYAGDTI 120	
DB	61 YVYSGASGIVKPEDMVDGGINNWSVLPLPSARLQAVKNSVAPAVYKSESKRYAGDTI 120	
QY	121 LCVRVLPFPEYSSOSAMIMPFPKIPFYSGSGNOFLKGLIDNTKTMKEIVSYSLCYET 180	
DB	121 LCVRVLPFPEYSSOSAMIMPFPKIPFYSGSGNOFLKGLIDNTKTMKEIVSYSLCYET 180	
QY	181 DIEVLPEDMNGMEIAYSMGTLKFKGADLIWNPNTIPINISRIKDDVNPPLASSKMR 240	
DB	181 DIEVLPEDMNGMEIAYSMGTLKFKGADLIWNPNTIPINISRIKDDVNPPLASSKMR 240	
QY	241 FFAFRVRSKSHSSVKKFIYVVDRLVLYOKLSVDSIDSDISEGVFKYENSSTESLRKTK 300	
DB	241 FFAFRVRSKSHSSVKKFIYVVDRLVLYOKLSVDSIDSDISEGVFKYENSSTESLRKTK 300	
QY	301 AHTEFKRVLKLRKISIAEGSFQNFYEKIESEKPEESSPKN 341	
DB	301 AHTEFKRVLKLRKISIAEGSFQNFYEKIESEKPEESSPKN 341	
RESULT 2		
AAV20146	AAV20146 standard; Protein; 335 AA.	
AC	AAV20146;	
XX		
PT	19-JUL-1999 (first entry)	
XX		
DE	B. burgdorferi antigenic protein, f07A.aa B8023.	
XX		
KW	Antigenic protein; vaccine; Lyme disease; infection; detection.	
XX		
OS	Borrelia burgdorferi.	
PN	W09859071-A1.	
XX		
PD	30-DEC-1998.	
XX		
PF	18-JUN-1998; 98MO-US12718.	
XX		
PR	03-SEP-1997; 97US-0057483.	
PR	20-JUN-1997; 97US-0050359.	
PR	22-JUN-1997; 97US-0053344.	

PR		22-JUL-1997;	97US-0053377.	
XX		(HUMA-) HUMAN GENOME SCI INC.		
PA		(MEDI-) MEDIMUNE INC.		
XX				
PI		Choi GH, Erwin AL, Hanson MS, Lathigra R;		
XX				
DR		WPI: 1999-189980/16.		
XX		N-P8DB; AAX61843.		
PT		New isolated Borrelia burgdorferi nucleic acids - used to develop		
PR		products for the diagnosis, prevention and treatment of diseases		
PT		caused by Borrelia, particularly Lyme disease		
XX				
PS		Claim 12; Page 214; 275pp; English.		
CC		This sequence represents a Borrelia burgdorferi (Bb) protein of the		
CC		invention, which is suitable for use in a vaccine. The Bb polypeptides		
CC		can be used in vaccines for eliciting protective antibodies to members of		
CC		the Borrelia genus, particularly for the use against Lyme disease in		
CC		humans and animals. They can be used for preventing or attenuating an		
CC		infection caused by a member of the Borrelia genus. The products can also		
CC		be used for detection of members of the Borrelia genus.		
XX				
SQ	Sequence	335 AA:		
	Query Match	95.7%: Score 1645; DB 20; Length 335;		
	Best Local Similarity	97.6%; Pred No. 1.3e-154;		
	Matches 327; Conservative	2; Mismatches 6; Indels 0; Gaps 0		
OY	7	STIFLLSTVLEFAOETDGLAEGSKRAPELVLDFAEALARDPSSRLDLTNYVDVYSGA 66		
Db	1	slflfllstvlfagetdglagsgkraepgelvidfaelardpsrldltlnydyvysga 60		
OY	67	GSIVRPEDMDVNDLGITNNMSVILLTPSARLDAAVYNVSVAAPAVYSESKRYAGDTIIIGRVL 126		
Db	61	sgivrpedmvdvdlglnmvslltpsarldayavnsvapavyseskrtyagdclilgyrvl 120		
OY	127	PFSYSOSSAMIPPEKIPRYPYSGESGNQFLGKLIDNITMKKEIKSVSYSLGYEIDEVLRF 186		
Db	121	fpsygsasamippekflypysgesgnqflgkglidnltmkkelksvyslygeidevlrf 180		
OY	187	EMNGMEIAYVSMGTILFKGMADLIWNPVNIPIISSRIIKDDVPNPLASSKKRRFKAFRV 246		
Db	181	edmnkmeyaxsmngclkfkfwadliwspnyipisssriikddvpnplasskkmrffkafrv 240		
OY	247	SKSHSKVYNLFIFYKYDLNLVKDKLSVSDISDSSEVFKVYTSGTESLRKLKAHETFK 306		
Db	241	skskshskgnlfifykydvrlvydklsvtsdsdsesvfkvysgtsgteslrklkaixctfk 300		
OY	307	RVLKUREKISIAGSFONFVEKIESCKPDESSPKN 341		
Db	301	xvllkirekismpegsfqnfvkekiesckpeesspkn 335		
	RESULT 3			
	AAY20147			
ID	AAY20147 standard; Protein; 315 AA.			
XX				
AC	AAY20147;			
XX				
DT	19-JUL-1999 (first entry)			
XX				
DE	B. burgdorferi antigenic protein, t07A.a			
XX				
KW	Antigenic protein; vaccine; Lyme disease; infection; detection.			
XX				
OS	Borrelia burgdorferi.			
XX				
PN	W09659071-A1.			
XX				
PD	30-DEC-1998.			



XX 18-JUN-1998; 98WO-US12718.  
XX  
XX 03-SEP-1997; 97US-0057483.  
PR 20-JUN-1997; 97US-0050359.  
PR 22-JUL-1997; 97US-0053444.  
PR 22-JUL-1997; 97US-0053377.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMUNE INC.  
XX  
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX  
XX WPI: 1999-189980/16.  
DR N-PSDB; AAX61844.  
XX  
XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
PT caused by Borrelia, particularly Lyme disease  
XX  
XX Claim 12; Page 214; 275pp; English.  
XX  
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus.  
XX  
SQ Sequence 315 AA;

Query Match 90.2%; Score 1551; DB 20; Length 315;  
Best Local Similarity 97.5%; Pred. No. 2.5e-145;  
Matches 307; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
QY 27 ESKRAPEPELVDFAEELARDSSRTLDITNVVDYVYSGASGVKREDMVVDGIMNSV 86  
DB 1 ESKRAPEPELVDFAEELARDSSRTLDITNVVDYVYSGASGVKREDMVVDGIMNSV 60  
QY 87 LTPSARLQAVYKNSVAVPAVVKSESKRYAGDTILGVRVLPFSQSSAMIMPFKIPRY 146  
DB 61 lltpsarldayvknsvavavvsksekryagdtllgvrllfpsysssamimppfipy 120  
QY 147 SEESNQIFGKGLIDINKTKEIKVSVSLGYEIDLEVLFEEDMNGMEYAVSMGTLEKGM 206  
DB 121 sgesngqifgkglidinktmekikvsvslgyeidlevlfednmngmeyaxsmgtlkfkgw 180  
QY 207 ADLIMNPNYININISRIIKDDVPNPLASSKMKRKAFFVSKSHSKVKNFTIYVVDLAV 266  
DB 181 adlwnpnnyipnissrllkddvpnpplasskknrfkafvskshskenflfivvkdlyv 240  
QY 267 LYDKLSVSDSIDSESVKVEYETSGTESLRKLKAHETFKRVLKREKISIAESQNFY 326  
DB 241 lydklsvdsidsdesvfkveysteslrklkahetfkrlkrekisiampegstqntfy 300  
QY 327 EKIESKEPESSPKN 341  
DB 301 ekiesekpeespkn 315  
RESULT 4  
ID AAR33279 standard; Protein; 320 AA.  
XX AAR33279;  
AC AAR33279;  
XX  
XX 16-JUL-1993 (first entry)  
DT  
XX 43 kD endoflagellum sheath protein.  
DE  
XX Endoflagellum; sheath protein; T. hyodysenteriae; core; antibody;

KW bacteriacide; 43 kD; vaccine; infection; swine dysentery.  
XX  
XX Treponema hyodysenteriae.  
OS  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..19  
FT */note="Signal peptide"*  
FT Protein 20..320  
FT */note="Mature protein"*  
XX  
XX EP534526-A.  
PN  
XX  
XX 31-MAR-1993.  
PD  
XX  
XX 14-SEP-1992; 92EP-0202796.  
XX  
XX 25-SEP-1991; 91EP-0202478.  
PR 24-JUL-1992; 92EP-0202273.  
XX  
XX (DUTN ) DUPHAR INT RES BV.  
PA  
XX  
XX Koopman MBH, Kusters JG;  
PI  
XX  
XX WPI: 1993-102665/13.  
DR N-PSDB; AAG38583.  
XX  
XX Vaccine to protect pigs against swine dysentery - comprises  
PT Treponema hyodysenteriae endo-flagellum sheath protein, applied  
PT orally or intranasally  
XX  
XX  
XX Claim 2; Page 21-22; 34pp; English.  
XX  
XX This sequence represents the endoflagellum sheath protein of T.  
CC hyodysenteriae. The endoflagellum consists of at least four  
CC proteins, this protein forms the sheath of the flagellum and three  
CC proteins, of molecular weights 37, 34 and 32 kD, make up its core.  
CC Antibodies raised against the sheath protein have been shown to be  
CC bactericidal for T. hyodysenteriae. The 43 kD sheath protein can  
CC be used in the production of a vaccine against infections such as  
CC swine dysentery.  
XX  
SQ Sequence 320 AA;

Query Match 18.9%; Score 324.5; DB 14; Length 320;  
Best Local Similarity 28.4%; Pred. No. 1e-23;  
Matches 89; Conservative 63; Mismatches 130; Indels 31; Gaps 8;  
QY 9 LPEFLSTVLFAPQETDGLAESKRAPEGELVLDFAELARDSSRTLDITNVVDYVYSGASC 68  
DB 4 ltvflstvlfaqetdglaeaakrapelgvldfaelardpsrtlditnvvdvyvgasc 43  
QY 69 IYKPEDMVVDLGIMNSVLLTPSARLQAVYKNSVAVPAVVKSESKRYAGDTILGVRVLP 128  
DB 44 thevypvaenlydnrvvwnlnesarltenrnsvyenvdskngngweagklygvrvhfp 103  
QY 129 SYSQSS-AMIMPFKIPFISGESGNQFL-GKGLIDINKTKEIKVSVSLGYEIDLEVL 186  
DB 104 laamnsyalvktvylemgygagdktyteggkgyldhvgelksisawvygtrnylfvnl 163  
QY 187 EDMNGMEYAVSMGTLEKFKGADLIMSNPNYIPNISRRIIKDDVPNPLASSKMKRFAFRV 246  
DB 164 qnefgelksypmgltvfyngvrgvwenreylpnrivdsavlre-plyprmlpsvkldslgf 222  
QY 247 SKSHSKVKNFTIYVVDLRYLYDKLSVSDSIDSESVKVEYETSGTESLRKLKAHETK 306  
DB 223 yrtkdktgsgdfltyvkdvtleydvvvvvdfeediddeatwqllk---tendrk-gates-- 276  
QY 307 RVLKLREKISIAE 319  
DB 277 --arlrgeaelrd 287

[illegible]

```

QY 101 SVMAVAVSKSKRRAGDTIIGVRLVLFPSYSSQSSAMIP-----PFKLPFSGEGGNFL 155
Db 71 mpqayrmagwqgkckearfiylveckfrrggnwldlptkgsqsyelpilgvsg----- 126
QY 156 GKGLDINIKTKKEIKRVSYSLGLEYIDLEVLFDPMNGMEYASMGTLKRGWADLIWSNP 215
Db 127 -----fdvwagagyslalealvrdctgrhltllgnldlfgwkhlsvspt 173
QY 216 YIPNISRRIINDVQVNYPLASSKMFKAFRVRKSHSSKVKKNPIFFVKDLRYLXDSLVS 275
Db 174 hipqts-----rylgseqhstfvgfrltspservdtly-----vffogfkala 217
QY 276 DSDID 280
Db 218 nmhid 222

RESULT 6
AAW81355
ID AAW81355 standard; Protein: 1657 AA.
XX
AC AAW81355;
XX
DT 16-APR-1999 (first entry)
XX
DE A. pleuropneumoniae AprxIV toxin serotype 1 protein sequence.
XX
KW AprxIV toxin; RTX toxin; attenuated microorganism; subunit vaccine;
XX A. pleuropneumoniae infection; diagnosis.
XX
OS Actinobacillus pleuropneumoniae.
XX
PN EP875574-A2.
XX
PD 04-NOV-1998.
XX
PF 08-APR-1998; 98EP-0201115.
XX
PR 10-APR-1997; 97EP-0201032.
XX
PA (ALKU ) AKZO NOBEL NV.
XX
PI Frey J, Segers RPAM;
XX
PI WPI: 1999-027956/03.
XX
DR N-PSDB: AAW68471.
XX
PT Attenuated live Actinobacillus pleuropneumoniae - and a nucleotide
PT sequence having promoter controlling expression of AprxIV gene, a
PT sub-unit vaccine containing pure AprxIV toxin, and an attenuated live
PT vaccine
XX
PS Example 1; Page 13-20; 56pp; English.
XX
XX
CC This sequence represents the AprxIV toxin serotype 1 isolated from
CC Actinobacillus pleuropneumoniae strain 4074 (serotype 1 reference
CC strain). The invention relates to an attenuated live Actinobacillus
CC pleuropneumoniae microorganism producing no functional AprxIV toxin. The
CC invention also provides a subunit vaccine for protecting an animal
CC against infection of A. pleuropneumoniae, the vaccine contains pure AprxIV
CC toxin and a pharmacologically allowable carrier. The pure AprxIV toxin can
CC be used for the diagnosis based on the presence of an antibody against
CC AprxIV for distinguishing A. pleuropneumoniae infection from A. suis.
CC N.B. This sequence was indexed from EP-875574, which is the first major
CC country equivalent to JP10290695.
XX
XX
Sequence 1657 AA;

```





28-JAN-1999.  
15-JUL-1998; 98WO-US14679.  
22-JUN-1998; 98US-0102322.  
17-JUL-1997; 97US-0896164.  
10-OCT-1997; 97US-0061599.  
10-OCT-1997; 97US-0061765.  
10-OCT-1997; 97US-0948705.  
11-OCT-1997; 97GB-0021697.  
(LUDW-) LUDWIG INST CANCER RES.  
Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,  
Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;  
Tureci O;  
WP1; 1999-132448/11.  
New isolated cancer associated nucleic acids and polypeptides -  
isolated using sera from cancer patients, used to develop products  
for the diagnosis, monitoring or treatment of cancers  
Disclosure; Page 728-730; 787pp; English.

```

Query Match Similarity      6.2%; Score 107; DB 20; Length 947;
Best Local Similarity      18.5%; Pred. No. 0.21;
Matches       76; Conservative    80; Mismatches   120; Indels   134; Gaps     20.

OY      1 MKRKKSLIFLFLSTVTLFAQETDGLAEGSKRAEPGEVLDFELARPDSST-RLDLTNVY 59
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      118 magdaleklfmqglsgmpgeeqvvyvkeriklktqgnlavssakeksspatekvifgqe 177
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      60 DTVSGASGIKVPEDMYVDLGINNMSVLTTPSARLOAQV-----NSVPAVAVKSESKRY 115
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      178 pavfkts-!splnvvgasvnsss---qtaaqytkgykrkdttpttsavkaase-- 230
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      116 AGDTLLGRVLPSPYSQSAMIMPPFK-----IPFVSGESGNOFLGKLIDNITKME 168
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      231 -----fspttks-valppkcmprnvp----dsqqy---nvvelvtkeq 272
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      169 IK-----VSYSIG-----YEI-----DLEVLFDINMGMEY- 194
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      273 lfhcsellkmlakhhfswapfgynpvdnaqlhnydvxknpmolglikemdqeyk 332
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      195 -AYSGTGLFKGMADLMGN-----PNVIPMISSRIKDQVNPVPLASKMKRFKAFRS 247
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
          -----faadvrlmfmcykynppdlevvltmaamqd-----VF 369
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      248 KSHSKSVKNFIPIYVDLRVLYDKLSVIDSIDSDSEGFVEYETSGTESRLIKAH----- 302
          ||| | | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      370 ehcfkip-lpevsmpcyiktat-----ettgrennessgsnsd 413
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      303 ---ETFRRVLKRERIS-----IAEGSQNFVEKIETSEKPEESSPK 340
          | ||| | | | : | : | : | : | : | : | : | : | : | : | : | : | :

```

```

Db      414 dsecdvrvrlaklqglkavhqqqlvlsqvprrlknkkkskkkkkkk 463

RESULT 11
ID      AAY19935
AC      AAY19935 standard; Protein; 1087 AA.
XX
XX      AAY19935;
XX
XX      19-JUL-1999 (first entry)
XX
XX      B. burgdorferi antigenic protein, t742.aa.
XX
XX      Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
XX      Borrelia burgdorferi.
XX
XX      MO9859071-A1.
XX
XX      30-DEC-1998.
XX
XX      18-JUN-1998; 98WO-US127718.
XX
XX      03-SEP-1997; 97US-0057483.
XX      20-JUN-1997; 97US-0050359.
XX      22-JUL-1997; 97US-0053344.
XX      22-JUL-1997; 97US-0053377.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (MEDI-) MEDIMMUNE INC.
XX
XX      Chol GH, Erwin AL, Hanson MS, Lathigra R;
XX
XX      WPI: 1999-189980/16.
XX      N-PSDB: AAX61632.
XX
XX      New isolated Borrelia burgdorferi nucleic acids - used to develop
XX      products for the diagnosis, prevention and treatment of diseases
XX      caused by Borrelia, particularly Lyme disease
XX
XX      Claim 12; Page 125; 275pp; English.
XX
XX      This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX      invention, which is suitable for use in a vaccine. The Bb polypeptides
XX      can be used in vaccines for eliciting protective antibodies to members of
XX      the Borrelia genus, particularly for the use against Lyme disease in
XX      humans and animals. They can be used for preventing or attenuating an
XX      infection caused by a member of the Borrelia genus. The products can also
XX      be used for detection of members of the Borrelia genus.
XX
XX      Sequence 1087 AA;

Query Match      6.1%; Score 104.5; DB 20; Length 1087;
Best Local Similarity 21.7%; Pred. No. 0.46;
Matches 68; Conservative 44; Mismatches 110; Indels 91; Gaps 12;

Oy      98 VKNSTVAAPVAVVSESKRRAGDTITLGVRLVFPYSOSSSAMIM-----PPRKIPRY 146
Db      14 VHHHSI: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      14 vhhsvi d---rnyk kys---vakllqdkypnedamlntlaelaanspfeskd 64

Oy      147 SEGSENOFLG--KGIDINIKT-----MKETKSVSYSLGY-EIDLEVLFPDMNGM 192
Db      65 qdsasngllldkikgg-dntktvnenfdiafmrlylksdgtstengsdrndvdyieded 123

Oy      193 EYAYMSGTLFEKGWADLWSNPNTYIPNISSRIIKDDVPNYPYPLASSKMRKAFRVSKSHS 252
Db      124 efkkskipeklk-----pntnpkeedqllgspnpklsvndqknlfnekiknlsg 174

Oy      253 KKKNIIFYVKDRLVLYDKLVSIDSDDISESYFKYVENS-----GTESLRKL----- 299
Db      175 knsenllndsgkleindqgnliskexsenellkcpdshkysnannctslkklssnsgqe 234

```

```

Oy 300 -----KAHETKRLVK-----LRKTSIA-EESFONF 325
Db 235 selspsqtlgklyrpyylikkelyladdintgrvltgknrlkellkglksnktgkv 294
Oy 326 VEKIESEKPEESS 338
   | | | : | : | |
Db 295 nellensknkeas 307

RESULT 12
AAV19934
AAV19934 standard; Protein: 1119 AA.
AAV19934;
AAV19934;
19-JUL-1999 (first entry)
B. burgdorferi antigenic protein, f742.aa.
Antigenic protein; vaccine; Lyme disease; infection; detection.
Borrelia burgdorferi.
OS
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PE 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI: 1999-189980/16.
DR N-PSDB: AAX61631.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 124-125; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 1119 AA;

Query Match 6.1%; Score 104.5; DB 20; Length 1119;
Best Local Similarity 21.7%; Pred. No. 0.48;
Matches 68; Conservative 44; Mismatches 110; Indels 91; Gaps 12.

Oy 98 YKNSVAVAPAVKSESKRRYAGDTIILGVRLFPSPYSOSSANIM-----PPKIRPY 146
   | | | | : : | : : | : | : | : |
Db 46 vnsuvid-----fnykks-----vakllgkypqnediamltlnlaeiansfskdl 96
   : | | | | : | | | | : | | | : | : | : | : |
Oy 147 SEESGNQFLG--KGLIDNIXT-----MKELKSVYSLGY-EIDLEVLFPEDMGNM 192
   : | | | | : | | | | : | | | : | : | : | : |
Db 97 grdsanqldkikgq-dnktknvnenfdiafnryikdstleenyssdrddvgiededis 155
   | : | : | : | : | : | : | : | : | : | : | : |
Oy 193 EAVYAMGTLEFKGMADLIWNSPNYIPNMISSRIITKIDVDPVYPLASSMRKTRKAFRVSKSHS 252
   | : | : | : | : | : | : | : | : | : | : | : |

```

Db 156 etkkskipeklk-----ptnupkveedjldqspmklsyvndqklnfleklnkslg 206

Qy 233 KVKNFIFVVKDLRVLYDKLSVSDSDIDSEVFVETSS-----GIESIRKL----- 299

Db 207 ksenenlndsqkiendkqntnlksksenenllktpdnskysnmnltslkksinsgqe 266

Qy 300 -----KHEETKRYLK-----LREKISIA-EGSPONF 325

Db 267 selspsstligklyrpsylkelyelloddntgrvtlgnrlkelkkglsnftqkv 326

Qy 326 VEKIESEKPEESS 338

Db 327 nellensknkeas.339

RESULT 13

AA682627

ID AAC62627 standard; Protein; 391 AA.

XX AAC62627;

XX AC

XX DT 03-SEP-2001 (first entry)

XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2348.

XX KM Staphylococcus epidermidis SRI strain; infection; diagnosis;

XX KW vaccination; endocarditis.

XX OS Staphylococcus epidermidis.

XX PN WO200134809-A2.

XX PD 17-MAY-2001.

XX PF 09-NOV-2000; 2000WO-US30782.

XX PR 09-NOV-1999; 99US-0164258.

XX PA (GLAX ) GLAXO GROUP LTD.

XX PI kimmerly WJ;

XX DR WPI; 2001-316495/33.

XX DR N-PSDB; AAH53477.

XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,

XX PT useful for vaccinating against infections, e.g. endocarditis -

XX PS

XX Claim 18; Page 621; 2188bp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAC81454 to AAC81320 from Staphylococcus epidermidis.

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the

CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to

CC AAH5090 represent specifically claimed S. epidermidis genomic DNA

CC polynucleotide sequences from the present invention. AAH5091 to

CC AAH5098 represent oligonucleotide sequences and primers which are used

CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4454 so even

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,

CC no sequences are present for SEQ ID NO:4455 to 4464.

XX

XX Sequence 391 AA;

Query Match 6.0% Score 102.5 DB 22 Length 391.  
Best Local Similarity 21.7% Pred.No.0.15;  
Matches 59; Conservative 39; Mismatches 91; Indels 83; Gaps 12

OY 117 GDITIGVRLFFSYSSQSSAMIMPFKIPFVSGSGNQFLGIDNIKTKKEIKVSYSYL 176  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Dd 45 getrlglnlh-----sealcip-----enqkyfpfngk-----datkgatlvnlygr 86  
OY 177 GYEIDLE-----VFEDNMGEAYASWGT-----LRFKWAD----- 208  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Dd 87 hydveldekdevelllygtkngl---valptcvvnpgeivllpppyldyagymladakp 143  
OY 209 --LIMSNIPIYN----ISSRIIKD-----DVNNYPLASKM-----RKRAFVSK 248  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 144 lplkslppylrpnwmttisakvlektkllyltypnmctslatqddfoeahrtikytltki 203  
OY 249 SHSSRVKNFIFFVKDLRVLVDKLSTVIDSDISESVFKVYETSG-----TESL 296  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 204 vndffysafgfaknpislasnkakvaaleifsls--kyymsgfrvgfavgnkkmlgal 261  
OY 297 RKLKAHEFRKRVLKLRKISIAEGSFQNVEK 328  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 262 kkyqcltnagmfgaqlqdaatyalnhydeflek 293

RESULT 14  
AAW98762  
ID AAW98762 standard; Protein: 450 AA.  
XX  
AC AAW98762:  
XX  
DT 31-MAR-1999 (first entry)  
XX  
DE H. pylori GHPD 1097 protein.  
XX  
KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;  
KM peptic ulcer disease.  
XX  
OS Helicobacter pylori.  
XX  
PN WO9843478-A1.  
PD 08-OCT-1998.  
XX  
PF 01-APR-1998; 98WO-US06371.  
XX  
PR 29-JUL-1997; 97US-0902615.  
PR 01-APR-1997; 97US-0833457.  
PR 24-JUN-1997; 97US-0881227.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
XX  
XJ Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
DR WPI: 1998-542293/46.  
DR N-PSDB; AAX14481.

New isolated Helicobacter polynucleotides - used to develop products for the diagnosis, prevention and treatment of Helicobacter infections and gastrointestinal diseases

Claim 8; Page 1640-1642; 2054pp: English.

This sequence represents a Helicobacter pylori GHPD protein of the invention. The polypeptides can be used for preventing or treating Helicobacter infections, and gastroduodenal diseases associated with these infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used for the production of antibodies. The products can also be used for detection and diagnosis.

50	Sequence	450 AA;
	Query Match	5.9%, Score 101.5; DB 19; Length 450;
	Best Local Similarity	23.5%; Pred. No. 0.23;
	Matches 63; Conservative	41; Mismatches 95; Indels 69; Gaps 16;
OY	95 QAYVNSVAAPAVVSESKR--YAGDTLIGVAVLPFSPYSOSSAMIM--PPEKIP---PVS 147	
DB	71 gafknkhhlkplmnestekhvllkgdmldalxllqsysekltmylloppntknefiy 130	
OY	148 GESGNOFLGKLIDIKMKKEIKSVSYSLGYEIDLEVFEEDANGMEVAYSMTLKFKWA 207	
DB	131 gddfs-----sneevelktldgyekeldy---lknlf-----gskchsgwl 168	
OY	208 DLINSPNVPPIISSRIIKD-----VPNPVLASSKm-----RFKAP-----RVS 247	
DB	169 sfmy--prll--lakdlilkdgvflfslidnecaqlkilcdelfgegnfvaclkwkkkq 224	
OY	248 KSHSSKV---KNPFYVKDLRLVLDKLSVDISIDSDPSRYKVFETGSTELRKIAHE 303	
DB	225 pslskvavlelyllvyakdfs-lldklgidavvsodsk----plintsnlskr----- 273	
OY	304 TEFRVLKLRKIS-IAEGSFQNFVEKIE 330	
DB	274 yfkkgirvkdsdnfiksgkyqnktmtie 301	
	RESULT 15	
	AAM01037	
ID	AAM01037 standard; Protein; 419 AA.	
AC	AAM01037;	
XX		
XX	AAW01037;	
XX		
DT	19-JAN-1997 (first entry)	
XX		
DE	Mycoplasma 46-48 kDa protective antigen.	
XX		
KW	Antigen: vaccine; mycoplasmal pneumonia; swine enzootic pneumonia;	
XX	diagnosis; antibody.	
XX		
OS	Mycoplasma hyopneumoniae strain Beaufort.	
XX		
FH	Key Location/Qualifiers	
FT	Misc-difference 70	
FT	/note= "codon 70 in the nucleotide sequence is	
FT	a stop codon"	
FT	Misc-difference 101	
FT	/note= "codon 101 in the nucleotide sequence is	
FT	a stop codon"	
FT	Misc-difference 254	
FT	/note= "codon 254 in the nucleotide sequence is	
FT	is a stop codon"	
XX		
PN	WO9628472-A1.	
XX		
PD	19-SEP-1996.	
XX		
PF	15-MAR-1996; 96WO-AU00149.	
XX		
PR	16-MAR-1995; 95AU-0001789.	
XX		
PA	(UYME ) UNIV MELBOURNE.	
XX		
PI	Doughty SW, Lee R, Walker J;	
XX		
DR	WPI. 1996-433763/43.	
XX	N-PSTD; AAT38241.	
XX		
PT	Putative protective antigens against Mycoplasma - used for the	
XX	detection, prevention or treatment of Mycoplasma infections, esp. M.	
XX	hyopneumoniae in swine	





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 00:00:39 ; Search time 1691.33 Seconds  
(Without alignments) 842.965 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663  
Sequence: 1 atgataacatttttccaanaa.....atactctgaagagaataatt 1663

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: N\_Geneseq\_1101.\*  
2: /SIDS2/gcgdata/geneseq/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/NA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/NA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/NA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/NA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/NA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/NA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/NA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/NA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/NA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/NA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/NA1999.DAT.\*  
22: /SIDS2/gcgdata/geneseq/NA2000.DAT.\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1637	98.4	1655	20	AA89520
2	989.6	59.5	1011	20	AA61843
3	924.6	55.6	945	20	AA61844
4	372.8	22.4	423	20	AA61740
5	372.8	22.4	474	20	AA61739
6	207	12.4	936	22	AA58252
7	207	12.4	936	22	AA58254
8	207	12.4	936	22	AA58257
9	207	12.4	936	22	AA58259
10	207	12.4	936	22	AA58262
11	207	12.4	938	22	AA58255

C	12	206	12.4	936	22	AA58252
C	13	206	12.4	936	22	AA58254
C	14	206	12.4	936	22	AA58257
C	15	206	12.4	936	22	AA58259
C	16	206	12.4	936	22	AA58262
C	17	206	12.4	938	22	AA58255
C	18	107	6.4	1140	14	AA038583
C	19	78	4.7	19124	18	AA72882
C	20	78	4.7	19124	21	AA28287
C	21	75.6	4.5	244	22	AA58238
C	22	72.6	4.4	244	22	AA58238
C	23	71.4	4.3	612	22	AA71471
C	24	67	4.0	6124	11	AA003568
C	25	64.8	3.9	3975	9	AA81157
C	26	64.8	3.9	3975	13	AA02299
C	27	63.6	3.8	723	22	AA71575
C	28	63	3.8	665	21	AA231996
C	29	61.6	3.7	335913	22	AA161371
C	30	61.6	3.7	335913	22	AA161372
C	31	60.4	3.6	3567	21	AA70117
C	32	58.8	3.5	1444	21	AA294422
C	33	58.4	3.5	4590	7	AA60472
C	34	58.2	3.5	4056	21	AA70225
C	35	58.2	3.5	10640	22	AA03729
C	36	57.8	3.5	1612	7	AA60392
C	37	57.8	3.5	2418	13	AA027866
C	38	57.6	3.5	2104	13	AA025273
C	39	57.6	3.5	5454	21	AA70236
C	40	57.2	3.4	9789	17	AA741852
C	41	56.8	3.4	605	17	AA731530
C	42	56.8	3.4	6124	11	AA003568
C	43	56.2	3.4	513445	22	AA161373
C	44	56	3.4	1251	21	AA70119
C	45	56	3.4	8920	15	AA062924

## ALIGNMENTS

RESULT 1	AA89520
ID	AA89520 standard; DNA; 1655 BP.
AC	AA89520;
XX	
DT	19-OCR-1999 (first entry)
XX	
DE	Nucleic acid sequence encoding the Flaa gene product.
XX	
KM	P37; Flaa; Lyme disease; antigen; assay; splirochete; ss.
XX	
OS	Synthetic.
XX	
OS	Borrelia burgdorferi
XX	
EH	Key
FT	CDS
FT	Location/Qualifiers
FT	473..1498
FT	/*tag- a
FT	/product= "P37/Flaa protein antigen"
FT	473..538
FT	/*tag- b
FT	539..1495
FT	mat_peptide
FT	/*tag- c
XX	
PN	W09935272-A1.
XX	
PD	15-JUL-1999.
XX	
XX	
PF	06-JAN-1999; 99WO-US00196.
XX	
PR	08-JAN-1998; 98US-0004395.
XX	
PA	(INMR ) BIOMERIEUX INC.
XX	

Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D1  
Oligonucleotide D2  
Oligonucleotide D2  
Oligonucleotide D1  
43 kd endoflagellu  
Plasmodium var-7 g  
Plasmodium var-7 p  
Oligonucleotide D1  
Oligonucleotide D1  
Human cervical can  
Sequence encoding  
Malaria-specific g  
SERP gene. Plasm  
Human cervical can  
Human apolipoprote  
Soybean 240017 reg  
Soybean 240017 reg  
Plasmodium falcipa  
Plasmodium falcipa  
Sequence encoding  
Plasmodium falcipa  
Sequence encoding  
P. falciparum GBP13  
Sequence encoding  
Plasmodium falcipa  
cDNA encoding plas  
Human 3' apolipopr  
Sequence encoding  
Soybean 318013 reg  
Plasmodium falcipa  
Carbamoyl-phosphat

Apple.

PI Gilmore RD, Johnson BJB;  
 XX WPI: 1999-430399/36.  
 DR P-PSDB; AAY28460.  
 XX  
 PT Recombinant FlaA/p37 protein antigen, useful in improved detection  
 of Borrelia burgdorferi  
 XX  
 PS Disclosure: Page 28; 34pp; English.  
 XX  
 CC The sequence encodes the recombinant protein antigen p37/FlaA. The  
 CC p37 protein product can be used in an assay for the detection of Borrelia  
 CC burgdorferi infection, the causative agent of Lyme disease. FlaA is  
 CC an outer sheath protein of the periplasmic flagella of B. burgdorferi.  
 CC FlaA is a prominent antigen in the early humoral immune response to  
 CC Borrelia burgdorferi infection, and significantly suitable for use in  
 CC improved serologic tests for exposure to Lyme disease spirochetes. The  
 CC detection of FlaA can augment the set of recombinant molecules that are  
 CC recognized early in the course of disease and contribute to improved  
 CC sensitivity of early testing for Lyme disease.  
 CC  
 XX Sequence 1655 BP; 560 A; 182 C; 308 G; 605 T; 0 other;

Query Match 98.4%; Score 1637; DB 20; Length 1655;  
 Best Local Similarity 99.5%; Pred. No. 3.2e-262;  
 Matches 1655; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 atgataatcttttcaaaaaggtttttatttcttctctagaaaggattgtgtcta 60  
 Db 1 atgataatcttttcaaaaaggtttttatttcttctctagaaaggattgtgtcta 60  
 QY 61 attaaagataatgaagagcaacaatttgaagaataattgaacatttgaataag 120  
 Db 61 attaaagataatgaagagcaacaatttgaagaataattgaacatttgaataag 120  
 QY 121 attctcttgaagaagaatgaggggttctcttgggttttattaaagtatcaaat 180  
 Db 121 attctcttgaagaagaatgaggggttctcttgggttttattaaagtatcaaat 180  
 QY 181 ttatatattttttttaaagaatgaggtttattatagtaaatcttcaagaagaatttatg 240  
 Db 181 ttatatattttttttaaagaatgaggtttattatagtaaatcttcaagaagaatttatg 240  
 QY 241 attctttaaataatgagatcaatcaatcttcttgaatgtaagaatcttttgcta 300  
 Db 241 attctttaaataatgagatcaatcaatcttcttgaatgtaagaatcttttgcta 300  
 QY 301 tgcctttaaataaacaatgagctttaaactatccctgttaagaataaagaatattatc 360  
 Db 301 tgcctttaaataaacaatgagctttaaactatccctgttaagaataaagaatattatc 360  
 QY 361 ttgttgaagggtgtatctttaaaggagcaaggttgataaatttactataaataat 420  
 Db 361 ttgttgaagggtgtatctttaaaggagcaaggttgataaatttactataaataat 420  
 QY 421 aattaaaaacggaataatttataaagattatataaagaggttggtttaacatgaaag 480  
 Db 421 aattaaaaacggaataatttataaagattatataaagaggttggtttaacatgaaag 480  
 QY 481 gaaagctaaaaagattatttattttttattatccactgtcttcttgcacaagactga 540  
 Db 481 gaaagctaaaaagattatttattttttattatccactgtcttcttgcacaagactga 540  
 QY 541 ttgattagcagaggtgtcttaaaaggcagagcctgtgagaattagcttagatttcgca 600  
 Db 541 ttgattagcagaggtgtcttaaaaggcagagcctgtgagaattagcttagatttcgca 600  
 QY 601 gcttgcaagagatccaagtccaactagactgacttaacaataatgcttgaatata 660  
 Db 601 gcttgcaagagatccaagtccaactagactgacttaacaataatgcttgaatata 660  
 QY 661 ttcggagcctctgtatgtttaagccgaagatagtgcttagatcttggataataa 720

Db 661 ttcggagcctctgtatgtttaagccgaagatagtgcttagatcttggataataa 720  
 QY 721 ttgagagcttttacttactctctctctgaaggttgagagcttaagtaaaatcagtgt 780  
 Db 721 ttgagagcttttacttactctctctctgaaggttgagagcttaagtaaaatcagtgt 780  
 QY 781 tgcgcgcgctgttctaagaatgagctcaaaaagtaacgagctgatactatttgggggt 840  
 Db 781 tgcgcgcgctgttctaagaatgagctcaaaaagtaacgagctgatactatttgggggt 840  
 QY 841 aagagtttcttccaagctatctccaatcatctgtatgataatgaccacattaaat 900  
 Db 841 aagagtttcttccaagctatctccaatcatctgtatgataatgaccacattaaat 900  
 QY 901 tccctttattcaggaggaatgagcaatcaattttaggaagaaggtcttataaat 960  
 Db 901 tccctttattcaggaggaatgagcaatcaattttaggaagaaggtcttataaat 960  
 QY 961 taaacacatgaagaatlaaggtatctgttataagtttagggatagataacttga 1020  
 Db 961 taaacacatgaagaatlaaggtatctgttataagtttagggatagataacttga 1020  
 QY 1021 ggtttatttgaagataatgaaatgagcaatgataatgcttataatggtaactt 1080  
 Db 1021 ggtttatttgaagataatgaaatgagcaatgataatgcttataatggtaactt 1080  
 QY 1081 taaaggttgagctgatttaatttggtaaatcccaatattcccaatataatcaccag 1140  
 Db 1081 taaaggttgagctgatttaatttggtaaatcccaatattcccaatataatcaccag 1140  
 QY 1141 aattataaagaagatgttccaatattccctctgtcttcaagaataaagatlaagc 1200  
 Db 1141 aattataaagaagatgttccaatattccctctgtcttcaagaataaagatlaagc 1200  
 QY 1201 tttaagattccaagctacacaggttcaaaaagttaaaaatttcaatcttataatga 1260  
 Db 1201 tttaagattccaagctacacaggttcaaaaagttaaaaatttcaatcttataatga 1260  
 QY 1261 tttaagattcttataatgataagctlaaggttccaatattcgaatcttgaatgagct 1320  
 Db 1261 tttaagattcttataatgataagctlaaggttccaatattcgaatcttgaatgagct 1320  
 QY 1321 tgcatttaaagttatagagctagcgaactgaatcccttcgttaaaatlaaaggcaca 1380  
 Db 1321 tgcatttaaagttatagagctagcgaactgaatcccttcgttaaaatlaaaggcaca 1380  
 QY 1381 aactttaaaaggtttaaagcttagaagaataattctatcgtcgaaggtcttcca 1440  
 Db 1381 aactttaaaaggtttaaagcttagaagaataattctatcgtcgaaggtcttcca 1440  
 QY 1441 aaacttgtgaagaagattgagaggtgaaaaacctgaagaatcatcccgaaaattag 1500  
 Db 1441 aaacttgtgaagaagattgagaggtgaaaaacctgaagaatcatcccgaaaattag 1500  
 QY 1501 ttaaatatcatgtaaagcaccctaaagggttgccttcttcttcttcttcttcttctt 1560  
 Db 1501 ttaaatatcatgtaaagcaccctaaagggttgccttcttcttcttcttcttcttctt 1560  
 QY 1561 taagaaatagatataatgaaatattagatttgaagaatgaaagctttagagattttt 1620  
 Db 1561 taagaaatagatataatgaaatattagatttgaagaatgaaagctttagagattttt 1620  
 QY 1621 tgaagaagctcaaatctgttgaatataatccttgaagaagaatt 1663  
 Db 1621 tgaagaagctcaaatctgttgaatataatccttgaagaagaatt 1663  
 QY 1663 tgaagaagctcaaatctgttgaatataatccttgaagaagaatt 1663  
 Db 1663 tgaagaagctcaaatctgttgaatataatccttgaagaagaatt 1663

RESULT 2  
 AAX61843  
 ID AAX61843 standard; DNA; 1011 BP.  
 XX  
 AC AAX61843;

Qy	849	tgttccaagactatctccaatcactcctgcgtatgatattgtgcaccattttaaattcctttt	908
Db	362	tggttccaagctctctccaatcactcctgcgtatgatattgtgcaccattttaaattcctttt	421
Oy	909	atccaaggagaaatgaggaacatcaatttttttaggcagaaggtctta ttgataacatlaaacca	968
Db	422	atccaaggagaaatgaggaacatcaatttttttaggcagaaggtctttttgataacatlaaacca	481
Oy	969	tgaaagaatthaagtgatcctgtttatagtttaaggtatgagatagaatccttgaggtttat	1038
Db	482	tgaaagaatthaagtgatcctgtttatagtttaaggtatgagatagaatccttgaggtttat	541
Oy	1029	ttgaaagatatgaaatgagatgagataatgactattcttatggtgactttaagtttaagagt	1088
Db	542	ttgaaagatatgaaatgagatgagataatgactattcttatggtgactttaagtttaagagt	601
Oy	1089	gggctgattcaatttggccaatccctaactatacttccataatcatccagaactatca	1148
Db	602	gggctgattcaatttggccaatccctaactatacttccataatcatccagaactatca	661
Oy	1149	aagaagatgttccaatatacctcctctgtctccaaglaaaa tgaagatttaaggtctttagag	1208
Db	662	aagaagatgttccaatatacctcctctgtctccaaglaaaa tgaagatttaaggtctttagag	721
Oy	1209	tttcaaaagtcacacagcttccaagaattcaaaaattctactctttatgttaagaagttaaag	1268
Db	722	tttcaaaagtcacacagcttccaagaagaatcttcaactctttatgttaagaagttaaag	781
Oy	1269	ttctcttatgaatgaatgaggttccaatagatctgatatctgaacggtgagctcgatca	1328
Db	782	ttctcttatgaatgaatgaggttccaatagatctgatatctgaacggtgagctcgatca	841
Oy	1329	aagtttatgagactagcgcgaactgaaatcccttcgtataatlaaaggacacagaaacttta	1388
Db	842	aagtttatgagactagcgcgaactgaaatcccttcgtataatlaaaggacacagaaacttta	901
Oy	1389	aaagaagctttaaagcttaagagaaaataattctatgcttgaagagcctcttccaaaacttg	1448
Db	902	aaagaagctttaaagcttaagagaaaataattctatgcttgaagagcctcttccaaaacttg	961
Oy	1449	tagaaaaagattgagagtgaaaaaaccttgaaagaalcatcttcgaaaaaatag	1498
Db	962	tagaaaaagattgagagtgaaaaaaccttgaaagaalcatcttcgaaaaaatag	1011
RESULT 3			
AAx61844			
ID	AAx61844	standard; DNA: 945 BP.	
XX	AAx61844:		
AC	AAx61844:		
XX	19-JUL-1999	(first entry)	
DT			
XX			
DE	B. burgdorferi antigenic protein coding sequence, t07A.nt BB023.		
XX			
KW	Antigenic protein; vaccine; Lyme disease; Infection; detection; ss.		
XX	Borrelia burgdorferi.		
OS			
XX	WO9859071-A1.		
PN			
XX	30-DEC-1998.		
PD			
XX	18-JUN-1998:	98MO-US12718.	
PF			
XX	03-SEP-1997:	97US-0057483.	
PR	20-JUN-1997:	97US-0050359.	
PR	22-JUL-1997:	97US-0053344.	
PR	22-JUL-1997:	97US-0053377.	
XX			
RA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(MEDI-) MEDIMMUNE INC.		

PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
 XX WPI; 1999-189980/16.  
 DR P-PSDB; AAY20147.  
 XX  
 PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop  
 products for the diagnosis, prevention and treatment of diseases  
 caused by *Borrelia*, particularly Lyme disease  
 PS  
 XX  
 PS Claim 1; Page 214; 275pp; English.  
 CC This sequence encodes a *Borrelia burgdorferi* (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the *Borrelia* genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the *Borrelia* genus. The products can also  
 CC be used for detection of members of the *Borrelia* genus.  
 XX  
 SQ Sequence 945 BP; 300 A; 128 C; 192 G; 319 T; 6 other;

Query Match 55.6%; Score 924.6; DB 20; Length 945;  
 Best Local Similarity 98.4%; Pred. No. 1.5e-144;  
 Matches 930; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 551 gaaggtctcaaaaggcagagcctgtagagatgcttgaatttcgcgagcttgcaga 610  
 DB 1 gagggtctcaaaaggcagagcctgtagagatgcttgaatttcgcgagcttgcaga 60  
 OY 611 gatccaagtccaactagacttgatcttaacaattatgtgattatatttcggcgct 670  
 DB 61 gatccaagtccaactagacttgatcttaacaattatgtgattatatttcggcgct 120  
 OY 671 tctggtattgttaagcgcgaagatatggtttagatcttgggtaataatttgaagcgt 730  
 DB 121 tctggtattgttaagcgcgaagatatggtttagatcttgggtaataatttgaagcgt 180  
 OY 731 ttacttaccctctcgaaggttgcaggtctgaactgaacttaaaatcagtttgcgcgcgt 790  
 DB 181 ttacttaccctctcgaaggttgcaggtctgaactgaacttaaaatcagtttgcgcgcgt 240  
 OY 791 gtgtttaagatgtagtcaaaaagtagtcagcagtgatatactatttggggtaagagtttg 850  
 DB 241 gtgtttaagatgtagtcaaaaagtagtcagcagtgatatactatttggggtaagagtttg 300  
 OY 851 ttccaagctattctcaatcatctgtatgatatgcacacatttaaatctctttat 910  
 DB 301 ttccaagctattctcaatcatctgtatgatatgcacacatttaaatctctttat 360  
 OY 911 tcaggggaagtgagcaataatttttagcgaaggtcttatatatacaattaaaccatg 970  
 DB 361 tcaggggaagtgagcaataatttttagcgaaggtcttatatatacaattaaaccatg 420  
 OY 971 aaagaataaagtagtctgttatagtttagggtagatagatcttgaggtttatt 1030  
 DB 421 aaagaataaagtagtctgttatagtttagggtagatagatcttgaggtttatt 480  
 OY 1031 gaagatatgaatgagcagtaatatgtcttatctatggtgtaactttaaagttgaagtg 1090  
 DB 481 gaagatatgaatgagcagtaatatgtcttatctatggtgtaactttaaagttgaagtg 540  
 OY 1091 gctgatttaattgtgcaaatcctaactatattccaatatataccacgaattatataa 1150  
 DB 541 gctgatttaattgtgcaaatcctaactatattccaatatataccacgaattatataa 600  
 OY 1151 gacagatgtccaaatatcctctgtctcaagtaaaatgagattaaagcttttagagt 1210  
 DB 601 gacagatgtccaaatatcctctgtctcaagtaaaatgagattaaagcttttagagt 660  
 OY 1211 tcaaatgcacacagtcacaaagttaaataatcattcatctttatgtagtaagattgaagt 1270  
 DB 661 tcaaatgcacacagtcacaaagttaaataatcattcatctttatgtagtaagattgaagt 720

OY 1271 cttatgataaagctaagtggtttcaatagatctgtatattgacagtgagtcgtatttaa 1330  
 DB 721 cttatgataaagtgagtggtttcaatagatctgtatattgacagtgagtcgtatttaa 780  
 OY 1331 gttatgagactagcggaaactgaatccctctgtaataatgaagcagacgaacttttaa 1390  
 DB 781 gttatgagactagcggaaactgaatccctctgtaataatgaagcagacgaacttttaa 840  
 OY 1391 agagttttaaagcttagagaaaaatttctatcgtcgaaggtctttccaactttgta 1450  
 DB 841 agagttttaaagcttagagaaaaatttctatcgtcgaaggtctttccaactttgta 900  
 OY 1451 gaaaagattgagatgagaaaaacctgaagaatcatctccgaaaaat 1495  
 DB 901 gaaaagattgagatgagaaaaacctgaagaatcatctccgaaaaat 945

RESULT 4  
 AAX61740  
 ID AAX61740 standard; DNA; 423 BP.  
 XX  
 AC AAX61740;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein coding sequence, t216.nt.  
 XX  
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

OS *Borrelia burgdorferi*.  
 XX  
 PN W09859071-A1.  
 XX

PD 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12718.  
 XX  
 PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 XX

PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
 XX  
 DR WPI; 1999-189980/16.  
 DR P-PSDB; AAY20043.  
 XX

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop  
 products for the diagnosis, prevention and treatment of diseases  
 caused by *Borrelia*, particularly Lyme disease  
 PS  
 XX

PS Claim 1; Page 171; 275pp; English.  
 XX

CC This sequence encodes a *Borrelia burgdorferi* (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the *Borrelia* genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the *Borrelia* genus. The products can also  
 CC be used for detection of members of the *Borrelia* genus.  
 CC  
 XX

SQ Sequence 423 BP; 138 A; 31 C; 71 G; 183 T; 0 other;

Query Match 22.4%; Score 372.8; DB 20; Length 423;  
 Best Local Similarity 97.3%; Pred. No. 2e-53;  
 Matches 393; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

OY 1 atgataatctttttcaaaaaggtttttatttcaatctagcaaggtattgtgcta 60

|||||  
Db 20 atgataatcttttccaaaaggcttttattcttccatcttagcaaggatctgtgcta 79  
Qy 61 atttaagataatttaagagatgatacaaaatttgaagaataattagatcctttaaaag 120  
Db 80 atttaagataatttaagagatgatacaaaatttgaagaataattagatcctttaaaag 139  
Qy 121 atttcttcttaaggaagaatgaggttcttcttgggttcttaataagtgatcaaat 180  
Db 140 atttcttcttaaggaagaatgaggttcttcttgggttcttaataagtgatcaaat 199  
Qy 181 tttaatactctttttaaagaatgaggttattatgtaaatctttcaaggaatttaag 240  
Db 200 tttaatactcttctttaaagaatgaggttattatgtaaatctttcaaggaatttaag 259  
Qy 241 attctttaaataatggtgata-----taataatctttgaatgataagtgataac 291  
Db 260 attctttaaataatggtgataataatgataatgataatctttgagtgataagtgataac 319  
Qy 292 ttttgcatagtctttaaataaacaatgcgctttaaacttctgtgataagataaaaga 351  
Db 320 ttttgcatagtctttaaataaacaatgcgctttaaacttctgtgataagataaaaga 379  
Qy 352 ttattatcttctgtgaagggtgatactttaaaggaagcaaatgtga 395  
Db 380 ttgtattcttctgtgaagggtgatactttaaaggaagcaaatgtga 423

## RESULT 5

AAK61739  
ID AAK61739 standard; DNA; 474 BP.

XX AAK61739;

XX 19-JUL-1999 (first entry)

XX B. burgdorferi antigenic protein coding sequence, f216.nt.

XX Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

XX Borrelia burgdorferi.

XX MO9859071-A1.

XX 30-DEC-1998.

XX 18-JUN-1998; 98MO-US12718.

XX 03-SEP-1997; 97US-0057483.

XX 20-JUN-1997; 97US-0050359.

XX 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (MEDI-) MEDIMMUNE INC.

XX Chai GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI; 1999-189980/16.

XX P-PDB; AAY20042.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop

XX products for the diagnosis, prevention and treatment of diseases

XX caused by Borrelia, particularly Lyme disease

XX Claim 1; Page 171; 275pp; English.

XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the

XX invention, which is suitable for use in a vaccine. The Bb polypeptides

XX can be used in vaccines for eliciting protective antibodies to members of

XX the Borrelia genus, particularly for the use against Lyme disease in

XX humans and animals. They can be used for preventing or attenuating an

XX infection caused by a member of the Borrelia genus. The products can also

CC be used for detection of members of the Borrelia genus.  
XX Sequence 474 BP; 145 A; 36 C; 85 G; 208 T; 0 other;

Query Match 22.4%; Score 372.8; DB 20; Length 474;

Best Local Similarity 97.3%; Pred. No. 2e-53; Matches 393; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

Qy 1 atgataatcttttccaaaaggcttttattcttccatcttagcaaggatctgtgcta 60  
Db 71 atgataatcttttccaaaaggcttttattcttccatcttagcaaggatctgtgcta 130  
Qy 61 atttaagataatttaagagatgatacaaaatttgaagaataattagatcctttaaaag 120  
Db 131 atttaagataatttaagagatgatacaaaatttgaagaataattagatcctttaaaag 190  
Qy 121 atttcttcttaaggaagaatgaggttcttcttgggttcttaataagtgatcaaat 180  
Db 191 atttcttcttaaggaagaatgaggttcttcttgggttcttaataagtgatcaaat 250  
Qy 181 tttaatactctttttaaagaatgaggttattatgtaaatctttcaaggaatttaag 240  
Db 251 tttaatactcttctttaaagaatgaggttattatgtaaatctttcaaggaatttaag 310  
Qy 241 attctttaaataatggtgata-----taataatctttgagtgataagtgataac 291  
Db 311 attctttaaataatggtgataataatgataatgataatctttgagtgataagtgataac 370  
Qy 292 ttttgcatagtctttaaataaacaatgcgctttaaacttctgtgataagataaaaga 351  
Db 371 ttttgcatagtctttaaataaacaatgcgctttaaacttctgtgataagataaaaga 430  
Qy 352 ttattatcttctgtgaagggtgatactttaaaggaagcaaatgtga 395  
Db 431 ttgtattcttctgtgaagggtgatactttaaaggaagcaaatgtga 474

## RESULT 6

AAF58252  
ID AAF58252 standard; DNA; 936 BP.

XX AAF58252;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D1835.

XX Election-transfer group; ETM; mismatch; genotyping;

XX gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000MO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing election-transfer group, useful as labels in

XX hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX a single surface -  
XX Example 6; Page 127; 159pp; English.





```
DT      24-APR-2001   (first entry)
XX
DE      Oligonucleotide D2004.
XX
KM      Electron-transfer group; ETM; mismatch; genotyping;
KW      gene expression; ss.
XX
OS      Synthetic.
XX
PN      WO200107665-A2.
XX
PD      01-FEB-2001.
XX
PE      26-JUL-2000; 2000WO-US20476.
PR      26-JUL-1999;    99US-0145695.
PP      17-MAR-2000; 2000US-0190259.
PA      (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI      Umek RM;
DR      WPT; 2001-159728/16.
XX
PS      Nucleic acids containing electron-transfer group, useful as labels in
PT      hybridization assays, e.g. for genotyping, allowing repeat analyses on
PR      a single surface -
PX
PY      Example 6; Page 128; 159pp; English.
ZZ
CC      The present invention relates to a composition comprising two nucleic
CC      acids each containing an electron-transfer group (ETM) having
CC      different redox potentials. The invention is used for electronic
CC      detection of nucleic acids, especially of substitutions (mismatches)
CC      and single-nucleotide polymorphisms, e.g. for genotyping,
CC      monitoring gene expression.
CX      Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;
XX
SQ
Query Match          12.4%; Score 207; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 5e-26;
Matches     9; Conservative 555; Mismatches 225; Indels    0; Gaps    0.
```

Db	427	.....
Qy	481	gaaagctaaagaatttattttttatbtacactgtccttttgctaagagactga
Db	487	.....
Qy	541	tggatagcagaaggctctaagggcgagccctggagaaattagtccttagattgcga
Db	547	.....
Qy	601	gcttcaagagatccaagtccaactgaacctacaattagtgtatgatata
Db	607	.....
Qy	661	ttcggcgccttcyglatgtttaagccggaagatatggttgtagacttggaataaa
Db	667	.....
Qy	721	ttgagagcttacttaactccttcgcaaggtgcaggtcttaagtaaattcagttgt
Db	727	.....
Qy	781	tgcgccgcgc 789
Db	787	wgccccccc 795
 RESULT 10		
ID	AAF58262	
XX	AAF58262 standard; DNA; 936 BP.	
AC	AAF58262;	
DT	24-APR-2001 (first entry)	
DE	Oligonucleotide D2007.	
XX		
KW	Electron-transfer group; ETM; mismatch; genotyping; gene expression; ss.	
OS	Synthetic.	
PN	WO200107665-A2.	
PD	01-FEB-2001.	
Pf	26-JUL-2000; 2000MO-US20476.	
PR	26-JUL-1999; 99US-0145695.	
PR	17-MAR-2000; 2000US-0190259.	
PPA	(CLIN-) CLINICAL MICRO SENSORS INC.	
PI	Umek RM;	
DR	WPI; 2001-159728/16.	
PT	Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface	
PS	Example 6; Page 128; 159pp; English.	
CC	The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.	
Sequence	936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;	



Query Match 12.4% Score 207; DB 22; Length 936;  
Best Local Similarity 1.1% Pred. No. 5e-26;  
Matches 9; Conservative 555; Mismatches 225; Indels 0; Gaps 0;

QY 1 atgataatcttttcaaaaaggttttatttcttcaacgaaggattgttcta 60  
DB 7 www. .... 66  
QY 61 atttaagatacttaagagatgaacaaatttgaagataattagacttttagtaaaag 120  
DB 67 www. .... 126  
QY 121 atttctttaggaagcaatgagggtttcttcttggttttattatcaatgatacaagat 180  
DB 127 www. .... 186  
QY 181 tttaatacttttcaagaatgaggttattatgtlaaacttcaagaagaatttatg 240  
DB 187 www. .... 246  
QY 241 attccttcaaatggttgaattaatcctttagttagagtcacacttttgccta 300  
DB 247 www. .... 306  
QY 301 tgccttaataaaaacaatgccttaactacccctgtaagataaaaagataatattc 360  
DB 307 www. .... 366  
QY 361 tctgtaaagggtgatacttaagcaagatgttaaatattacttaataaataat 420  
DB 367 www. .... 426  
QY 421 aattaaaaacgaaaatttataaagattatataaaggagtgtgtacatagaag 480  
DB 427 www. .... 486  
QY 481 gaaagctaaagatttattttttattatccactgtcttcttgccaagaagactga 540  
DB 487 www. .... 546  
QY 541 tgcattagcagaagggttcttaaaaggcagagcctgaggaattagctttagatttcgca 600  
DB 547 www. .... 606  
QY 601 gcttgaagagatccaagtccaacttgaacttatacaaatatgttatagtata 660  
DB 607 www. .... 666  
QY 661 ttcggagcgtctcgtatcgttaagccggaagatacgtgtgataccttgggataataa 720  
DB 667 www. .... 726  
QY 721 ttgagagcttactactcctctgcaaggtgcagccttaccgttcaaaattcagttgt 780  
DB 727 www. .... 786  
QY 781 tgcgcgcgc 789  
DB 787 wgccccccc 795

RESULT 11  
AAFS8255  
ID AAF58255 standard; DNA; 938 BP.  
XX  
AC AAF58255;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Oligonucleotide D1876.  
XX  
KW Electron-transfer group; ETW; mismatch; genotyping;

KW gene expression; ss.  
XX  
OS Synthetic.  
XX  
PN WO200107665-A2.  
XX  
PD 01-FEB-2001.  
XX  
PF 26-JUL-2000; 2000MO-US20476.  
XX  
PR 26-JUL-1999; 99US-0145695.  
XX  
PR 17-MAR-2000; 2000US-0190259.  
XX  
PA (CLIN-) CLINICAL MICRO SENSORS INC.  
XX  
PI umek RM;  
XX  
DR WPI; 2001-159728/16.  
XX  
PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface  
XX  
PS Example 6; Page 127; 159pp; English.  
XX  
CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETW) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 12.4% Score 207; DB 22; Length 938;  
Best Local Similarity 1.1% Pred. No. 5e-26;  
Matches 9; Conservative 555; Mismatches 225; Indels 0; Gaps 0;

QY 1 atgataatcttttcaaaaaggttttatttcttcaacgaaggattgttcta 60  
DB 7 www. .... 66  
QY 61 atttaagatacttaagagatgaacaaatttgaagataattagacttttagtaaaag 120  
DB 67 www. .... 126  
QY 121 atttctttaggaagcaatgagggtttcttcttggttttattatcaatgatacaagat 180  
DB 127 www. .... 186  
QY 181 tttaatacttttcaagaatgaggttattatgtlaaacttcaagaagaatttatg 240  
DB 187 www. .... 246  
QY 241 attccttcaaatggttgaattaatcctttagttagagtcacacttttgccta 300  
DB 247 www. .... 306  
QY 301 tgccttaataaaaacaatgccttcaactcctgtaagataaaaagataatattc 360  
DB 307 www. .... 366  
QY 361 tctgtaaagggtgatacttaagcaagatgttaaatattacttaataaataat 420  
DB 367 www. .... 426  
QY 421 aattaaaaacgaaaatttataaagattatataaaggagtgtttagactgaag 480  
DB 427 www. .... 486  
QY 481 gaaagctaaagatttatttttattatccactgtcttcttgccaagaagactga 540  
DB 487 www. .... 546

Accession	Sequence	Length
Db	487	546
Oy	541	600
Db	547	606
Oy	601	660
Db	607	666
Oy	661	720
Db	667	726
Oy	721	780
Db	727	786
Oy	781	
Db	787	

RESULT	12
AAAF58252/c	
ID	AAAF58252 standard; DNA; 936 BP.

AC AAF58252;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1835.

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN W0200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476

PR 26-JUL-1999; 99US-0145695

XX

XX

XX

XX

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX

XX

CC acids each containing an electron-transfer group (ETM) having

CC detection of nucleic acids, especially of substitutions (mismatches)

CC monitoring gene expression.

Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match	12.4%	Score 206	DB: 22	Length 936
Best Local Similarity	0.68	Pred. No. 7.4e-26		
Matches	5	Conservative	219	Indels 0
				Gaps 0

[illegible]

RESULT	ID
13	AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

Electron-transfer group; ETM; mismatch; genotyping;

XX  
XX  
XX

XX 9

XX

XX 5

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

XX 26-JUL-1999: 99US-0145695.  
PR 17-MAR-2000: 2000US-0190259.  
XX (CLIN-) CLINICAL MICRO SENSORS INC.  
XX Umek RM:  
XX WPI: 2001-159728/16.  
DR  
XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
XX Example 6: Page 127; 159pp; English.  
PS  
XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 12.4%; Score 206; DB 22; Length 936;  
Best Local Similarity 0.6%; Pred. No. 7.4e-26;  
Matches 5; Conservative 554; Mismatches 219; Indels 0; Gaps 0;

Qy 1 atgataatctttccaaaaggcttttatttcattcctagcaaggagttgtgcta 60  
Db 778 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 719  
Qy 61 attaagataatgaagaatgaacaaattgaaagataattagacttttagtaaa 120  
Db 718 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 659  
Qy 121 attctctttagaagaagctgaggttcttcttgggttttataagatcaaat 180  
Db 658 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 599  
Qy 181 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 240  
Db 598 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 539  
Qy 241 attcttataatgaatgaatgaatcttctgagttaagatcaatctttgcta 300  
Db 538 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 479  
Qy 301 tgccttataaataaagaatgcgttcaactcctgtaagataaaagataatac 360  
Db 478 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 419  
Qy 361 tgcgtgaaggctgatacttaaaaggcaagatgataaatttaataataaat 420  
Db 418 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 359  
Qy 421 aattaaaaacgaataattataaagattatataaagagtggtttacatagaag 480  
Db 358 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 289  
Qy 481 gaaagctaaagattatttttttattatcactgttcttcttgcataagagctga 540  
Db 298 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 239  
Qy 541 tgcgtgaaggctgatacttaaaaggcaagcctgagaaattagcttagatttcgca 600  
Db 238 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 179  
Qy 601 gcttgcaagatgacaaatgtaactgactgacttacaataatggttataatgata 660  
Db 178 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 119

Qy 661 ttccggcgcttctgtagttgaagccgaagataigtgttagacttgggataaa 720  
Db 118 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 59  
Qy 721 ttggaagctttactactcctctcgaaggttcgaagcttacaataatcaat 778  
Db 58 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 1

RESULT 14  
ID AAF58257/c  
AAF58257 standard; DNA; 936 BP.

AC AAF58257;  
XX  
XX 24-APR-2001 (first entry)  
DT  
XX  
XX Oligonucleotide D1954.  
DE

XX Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
XX Synthetic.

XX MO200107665-A2.  
XX  
XX 01-FEB-2001.  
PD

XX 26-JUL-2000; 2000WO-US20476.  
PF  
XX

XX 26-JUL-1999: 99US-0145695.  
PR 17-MAR-2000: 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.  
XX

XX Umek RM:  
XX

XX WPI: 2001-159728/16.  
XX

XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX

PS Example 6: Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX

XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 12.4%; Score 206; DB 22; Length 936;  
Best Local Similarity 0.6%; Pred. No. 7.4e-26;  
Matches 5; Conservative 554; Mismatches 219; Indels 0; Gaps 0;

Qy 1 atgataatctttccaaaaggcttttatttcattcctagcaaggagttgtgcta 60  
Db 778 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 719  
Qy 61 attaagataatgaagaatgaacaaattgaaagataattagacttttagtaaa 120  
Db 718 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 659  
Qy 121 attctctttagaagaagctgaggttcttcttgggttttataagatcaaat 180  
Db 658 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 599  
Qy 181 ttttataatcttttcaaaaggcttttatttcattcctagcaaggagttgtgcta 240

```
DB 598 WWWW... 539
QY 241 attccttaataatgataatgaatccttgatgataagtcacatttgcta 300
DB 538 WWWW... 479
QY 301 tgccttaataaacaatgcgttaactcctcgtaagataaagaattatc 360
DB 478 WWWW... 419
QY 361 ttgtgaagggtgatacttaaaaggcagaagtgaataaataatttactaataaat 420
DB 418 WWWW... 359
QY 421 aattaaaaacgaataatttataaagaattatataaagggttggttacaagaag 480
DB 358 WWWW... 299
QY 481 gaaagctaaagtatttatttttttattatcacactgtcttctgctcaagagactga 540
DB 238 WWWW... 239
QY 541 tggattagcagaagggtcttaaaaggcagcctgagagataagcttagatttgcga 600
DB 238 WWWW... 179
QY 601 gcttgaagagatccaagtccaactagacttgatccttaacaatttgatgata 660
DB 178 WWWW... 119
QY 661 ttcggcgcttcgttatgttaagccggaagatagtgttgatgacttggaataaa 720
DB 118 WWWW... 59
QY 721 ttgagcgcttactactcctctgcgaaggtgcaggttactgtaaaattcagtt 778
DB 58 WWWW... 1

RESULT 15
AAFS8259/C
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
```

```
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 12.4%; Score 206; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 7 4e-26;
Matches 5; Conservative 554; Mismatches 219; Indels 0; Gaps 0;

QY 1 atgataatcttttcaaaaagggttttatttcttctagcaaggagatttgcta 60
DB 778 WWWW... 719
QY 61 attaagatattgaagatgaacaaattgaaagataattagatctttagtaaaag 120
DB 718 WWWW... 659
QY 121 atttctttaggaagcaatgaggggttcttcttgggtttattatgaatgaat 180
DB 658 WWWW... 599
QY 181 tttaatactttttaaagaatgagttatattgaatccttcaagagatttag 240
DB 598 WWWW... 539
QY 241 attccttaataatggtatbaatgaatccttgatgataagtcacatttgcta 300
DB 538 WWWW... 479
QY 301 tgccttaataaacaatgcgttaactcctcgtaagataaagaattatc 360
DB 478 WWWW... 419
QY 361 ttgtgaagggtgatacttaaaaggcagaagtgaataaataatttactaataaat 420
DB 418 WWWW... 359
QY 421 aattaaaaacgaataatttataaagaattatataaagggttggttacaagaag 480
DB 358 WWWW... 299
QY 481 gaaagctaaagtatttatttttttattatcacactgtcttctgctcaagagactga 540
DB 238 WWWW... 239
QY 541 tggattagcagaagggtcttaaaaggcagcctgagagataagcttagatttgcga 600
DB 238 WWWW... 179
QY 601 gcttgaagagatccaagtccaactagacttgatccttaacaatttgatgata 660
DB 178 WWWW... 119
QY 661 ttcggcgcttcgttatgttaagccggaagatagtgttgatgacttggaataaa 720
DB 118 WWWW... 59
QY 721 ttgagcgcttactactcctctgcgaaggtgcaggttactgtaaaattcagtt 778
DB 58 WWWW... 1
```

Mon Mar 25 16:08:02 2002

us-09-004-395-1.rng

Page 13



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 01:57:36 ; Search time 1691.33 Seconds  
(Without alignments)  
10.645 Million cell updates/sec

Title: US-09-004-395-3

Sequence: 1 atgaaaggaagcttaagt 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq.1101.\*

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	20	AXX89521
2	21	100.0	1655	20	AXX89520
3	17.4	82.9	626	21	AAC49540
4	17.4	82.9	627	21	AAC50183
5	17.4	82.9	678	21	AAC34709
6	16.8	80.0	383	22	AA135657
7	16.8	80.0	383	22	AA104115
8	16.8	80.0	407	22	AA166213
9	16.8	80.0	476	22	AA111559
10	16.8	80.0	476	22	AA132844
11	16.8	80.0	476	22	AA101480

12	16.8	80.0	2569	20	AXX40071	Colon cancer assoc
13	16.4	78.1	232	21	AA229782	Purative PMG-1 fro
14	16.4	78.1	346	22	AAH83805	Human ovarian tumo
15	16.4	78.1	346	22	AAH83806	Human ovarian tumo
16	16.4	78.1	512	22	AAH12296	Human cDNA clone (
17	16.4	78.1	525	22	AAH83296	Human ovarian tumo
18	16.4	78.1	683	22	AA117578	Human breast cance
19	16.4	78.1	687	22	AA144891	Human breast cance
20	16.4	78.1	760	21	AA142519	Arabidopsis thalia
21	16.4	78.1	769	20	AA197936	Human secreted pro
22	16.4	78.1	803	21	AA151986	Arabidopsis thalia
23	16.4	78.1	1257	21	AA175110	cDNA encoding a hu
24	16.4	78.1	1257	21	AA175114	Human PRO1550 cDNA
25	16.4	78.1	1257	22	AA192118	Human PRO1550 cDNA
26	16.4	78.1	1257	22	AA192118	Human PRO1550 cDNA
27	16.4	78.1	1286	21	AA164536	Probe #60 used in
28	16.4	78.1	1288	20	AA184525	Human secreted pro
29	16.4	78.1	1295	22	AA199687	Human protein enco
30	16.4	78.1	1342	21	AA126424	Human secreted pro
31	16.4	78.1	1619	21	AA129780	Sequence upstream
32	16.4	78.1	2002	22	AA114767	Human cDNA sequenc
33	16.4	78.1	130480	22	AA125833	R. marinus bacteri
34	16.2	77.1	26	21	AA170952	Single nucleotide
35	16.2	77.1	26	21	AA170955	Single nucleotide
36	16.2	77.1	26	21	AA170958	Single nucleotide
37	16.2	77.1	96	21	AA123158	Human secreted pro
38	16.2	77.1	238	21	AA170953	Single nucleotide
39	16.2	77.1	238	21	AA170956	Single nucleotide
40	16.2	77.1	238	21	AA170959	Single nucleotide
41	16.2	77.1	332	22	AA172390	Human cervical can
42	16.2	77.1	337	22	AA170860	Human cervical can
43	16.2	77.1	365	22	AA169505	Human cervical can
44	16.2	77.1	480	20	AA113749	Enterococcus faeca
45	16.2	77.1	575	22	AA172968	Human cervical can

# ALIGNMENTS

RESULT 1	
ID AAX89521	standard; DNA: 21 BP.
AC AAX89521:	
XX	
XX	19-OCT-1999 (first entry)
XX	
XX	PCR primer F1 for amplification of FlaA.
XX	
XX	P37; FlaA; Lyme disease; antigen; assay; PCR primer; F1 construct; ss.
XX	
XX	Synthetic.
XX	Borrelia burgdorferi.
XX	
XX	W09935272-A1.
XX	
XX	15-JUL-1999.
XX	
XX	06-JAN-1999; 99WO-US00196.
XX	
XX	08-JAN-1998; 98US-0004395.
XX	
XX	(INMR ) BIOMERIEUX INC.
XX	
XX	Gilmore RD, Johnson BJB;
XX	
XX	WPI; 1999-43039/36.
XX	
XX	Recombinant FlaA/P37 protein antigen, useful in improved detection
XX	of Borrelia burgdorferi
XX	
XX	Example 2; Page 11; 34pp; English.

CC The sequence is a PCR primer denoted F1, used to create the F1  
CC construct, which consists of the entire coding sequence of P37  
CC (AA89520). The P37 protein product can be used as an assay for the  
CC detection of Borrelia burgdorferi infection, the causative agent of Lyme  
CC disease. Flaa is an outer sheath protein of the periplasmic flagella of  
CC B.burgdorferi. Flaa is a prominent antigen in the early humoral immune  
CC response to Borrelia burgdorferi infection, and is significantly suitable  
CC for use in improved serologic tests for exposure to Lyme disease  
CC spirochetes. The detection of Flaa can augment the set of recombinant  
CC molecules that are recognized early in the course of disease and  
CC contribute to improved sensitivity of early testing for Lyme disease.  
CC  
SQ Sequence 21 BP; 12 A; 1 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 21;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaagaagaaagctaaagt 21  
|||||  
DB 1 atgaagaagaaagctaaagt 21

RESULT 2  
AA89520  
ID AA89520 standard; DNA; 1655 BP.

XX AA89520;

DT 19-OCT-1999 (first entry)

DE Nucleic acid sequence encoding the Flaa gene product.

XX P37; Flaa; Lyme disease; antigen; assay; spirochete; ss.

OS Synthetic.

OS Borrelia burgdorferi

FM Key Location/Qualifiers

FT CDS 473..1498  
/\*tag= a

FT sig\_peptide 473..538  
/\*tag= b

FT mat\_peptide 539..1495  
/\*tag= c

FT WO9935272-A1.

PD 15-JUL-1999.

PF 06-JAN-1999; 99WO-US00196.

PR 08-JAN-1998; 98US-0004395.

PA (INMR ) BIOMERIEUX INC.

PI Gilmore RD, Johnson BJB;

XX WPI; 1999-430399/36.

DR P-PSDB; AAY28460.

PT Recombinant Flaa/P37 protein antigen, useful in improved detection

PS of Borrelia burgdorferi

XX Disclosure; Page 28; 34pp; English.

CC The sequence encodes the recombinant protein antigen P37/Flaa. The  
CC P37 protein product can be used in an assay for the detection of Borrelia  
CC burgdorferi infection, the causative agent of Lyme disease. Flaa is  
CC an outer sheath protein of the periplasmic flagella of B. burgdorferi.  
CC Flaa is a prominent antigen in the early humoral immune response to

CC Borrelia burgdorferi infection, and significantly suitable for use in  
CC improved serologic tests for exposure to Lyme disease spirochetes. The  
CC detection of Flaa can augment the set of recombinant molecules that are  
CC recognized early in the course of disease and contribute to improved  
CC sensitivity of early testing for Lyme disease.  
CC  
SQ Sequence 1655 BP; 560 A; 182 C; 308 G; 605 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaagaagaaagctaaagt 21  
|||||  
DB 473 atgaagaagaaagctaaagt 493

RESULT 3  
AAC49540/C  
ID AAC49540 standard; DNA; 626 BP.

XX AAC49540;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 61536.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PM EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 25-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0134256.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.



PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.  
Query Match 82.9%; Score 17.4; DB 21; Length 626;  
Best Local Similarity 94.7%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 gaaaaggaagctaaagt 21  
|||||||  
Db 505 GAAAGGAAATGCTAAAGT 487  
RESULT 4  
AAC50183/c  
ID AAC50183 standard; DNA; 627 BP.  
XX  
AC AAC50183;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 63886.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
OS  
EN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135628.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143542.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0144005.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0158765.  
PR 07-OCT-1999; 99US-0158829.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 82.9%; Score 17.4; DB 21; Length 627;  
Best Local Similarity 94.7%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 3 gaaaggaagctaaagt 21

DB 507 GAAAGCATGCTAAAGT 489  
|||||||  
RESULT 5  
AAC34709/c  
ID AAC34709 standard; DNA; 678 BP.  
XX  
XX AAC34709;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7604.  
XX  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126764.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.

```
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147122.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160774.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.9%; Score 17.4; DB 21; Length 678;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 6
AAI35657
ID AAI35657 standard; DNA; 383 BP.
```

```
XX AC AA135657;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #4343 used to measure gene expression in human placenta sample.
XX KW Probe: microarray; human; placenta; antenatal diagnosis;
XX KM genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX DT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX CC Claim 25; SEQ ID NO 4343; 654pp; English.
XX PS The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 383 BP; 127 A; 69 C; 69 G; 118 T; 0 other;

Query Match      80.0%; Score 16.8; DB 22; Length 383;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgaagaaggaagctaaagt 21
   ||||||| |||||||
DB 145 tgaagaagtagctaaagt 164

RESULT 7
AA104115
ID AA104115 standard; DNA; 383 BP.
XX AC AA104115;
XX DT 09-OCT-2001 (first entry)
XX DE Probe #4106 used to measure gene expression in human breast sample.
XX KW Probe: human; breast disease; breast cancer; development disorder; ss;
XX KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
```

```
XX XX 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX DT Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast -
XX CC Claim 25; SEQ ID NO 4106; 322pp; English.
XX PS The present invention relates to novel single exon nucleic acid probes.
XX CC The present sequence is one such probe. The probes are useful for
XX CC measuring human gene expression in a human breast sample, where the probe
XX CC hybridises at high stringency to a nucleic acid expressed in the human
XX CC breast. The probes are useful for predicting, diagnosing, grading,
XX CC staging, monitoring and prognosing diseases of the human breast,
XX CC particularly those diseases with polygenic aetiology. The diseases
XX CC include: breast cancer, disorders of development, inflammatory diseases
XX CC of the breast, fibrocystic changes, proliferative breast disease and
XX CC non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcl_sequences.
XX SQ Sequence 383 BP; 127 A; 69 C; 69 G; 118 T; 0 other;

Query Match      80.0%; Score 16.8; DB 22; Length 383;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgaagaaggaagctaaagt 21
   ||||||| |||||||
DB 145 tgaagaagtagctaaagt 164

RESULT 8
AA166213/C
ID AA166213 standard; cDNA; 407 BP.
XX AC AA166213;
XX DT 09-APR-2001 (first entry)
XX DE Novel human polynucleotide, SEQ ID NO: 1969.
XX KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
XX KM breast cancer; lung cancer; cancer detection; ss.
XX OS Homo sapiens.
XX PN WO200102568-A2.
XX PD 11-JAN-2001.
XX PR 30-JUN-2000; 2000WO-US18374.
XX PR 02-JUL-1999; 99US-0142310.
XX PR 02-JUL-1999; 99US-0142311.
XX PA (CHIR ) CHIRON CORP.
```

PA (HYSE-) HYSEQ INC.  
XX  
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A,  
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lanson G, Dimaac R,  
PI Cirtenjakov R, Dimaac S, Dickson M, Labat I, Leshkowitz D;  
PI Kita D, Garcia V, Jones LW, Strache-Crain B;  
DR WPI: 2001-091805/10.  
XX  
XX Library of polynucleotides for diagnosing a cancerous state of a  
PT mammalian cell and detecting cancer, particularly of the colon or  
PT prostate, comprises 3351 human polynucleotide sequences -  
XX  
PS Claim 9; Page 828; 1046pp; English.  
XX  
CC The present sequence is one of 3351 sequences in a library of human  
CC polynucleotides. The library is used to detect differentially expressed  
CC genes correlated with a cancerous state of a mammalian cell and can  
CC detect colon, prostate, breast and lung cancer. The library can be used  
CC to produce probes for detection of mRNA and to produce additional copies  
CC of the polynucleotides. The probes can be used for chromosome mapping of  
CC the polynucleotide and for detection of transcription levels. Ribozymes  
CC or antisense oligonucleotides can be generated. The polynucleotides and  
CC their gene products are used as genetic or biochemical markers (e.g. in  
CC blood or tissues) that will detect the earliest changes along the  
CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
CC preventive interventions. The polynucleotides, polypeptides and  
CC antibodies against them can be used in pharmaceutical compositions to  
CC treat the cancers and proliferative disorders such as neoplasia,  
CC dysplasia and hyperplasia.  
XX  
SQ Sequence 407 BP; 109 A; 62 C; 87 G; 149 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 407;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tgaagaagaagctaaagt 21  
||||| |||||||  
DB 52 TCAAAATATAAGCTAAAGT 33

RESULT 9  
AI11559  
ID AI11559 standard; DNA: 476 BP.  
XX  
AC AA11559;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Probe #1492 for gene expression analysis in human cervical cell sample.  
XX  
KM Probe; human; microarray; gene expression; cervical epithelial cell;  
KM cervical cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00670.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI: 2001-488901/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
XX  
PS Claim 25; SEQ ID No 1492; 487pp; English.  
XX  
CC The present invention relates to human single exon nucleic acid probes  
CC (SENP). The present sequence is one such probe. The SENPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 476 BP; 154 A; 79 C; 81 G; 162 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 476;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tgaagaagaagctaaagt 21  
||||| |||||||  
DB 159 tgaagaagtagctaaagt 178

RESULT 10  
AI132844  
ID AI132844 standard; DNA: 476 BP.  
XX  
AC AA132844;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Probe #1530 used to measure gene expression in human placenta sample.  
XX  
KM Probe; microarray; human; placenta; antenatal diagnosis;  
KM genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00663.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI: 2001-48897/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
XX  
SQ Claim 25; SEQ ID No 1530; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.

XX  
SQ Sequence 476 BP; 154 A; 79 C; 81 G; 162 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 476;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tgaagaaggaagctaaagt 21  
||||||| |  
DB 159 tgaagaagctagctaaagt 178

## RESULT 11

ID AAI01480 standard; DNA; 476 BP.

AC AAI01480;

DT 09-OCT-2001 (first entry)

DE Probe #1471 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;  
KW Inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.

PF 29-JAN-2001; 2001WO-US00661.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632365.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-476286/51.

PT Novel single exon nucleic acid probe used to measuring gene expression

PT in a human breast -

PS Claim 25; SEQ ID No 1471; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.

CC The present sequence is one such probe. The probes are useful for

CC measuring human gene expression in a human breast sample, where the probe

CC hybridises at high stringency to a nucleic acid expressed in the human

CC breast. The probes are useful for predicting, diagnosing, grading,

CC staging, monitoring and prognosing diseases of the human breast,

CC particularly those diseases with polygenic aetiology. The diseases

CC include: breast cancer, disorders of development, inflammatory diseases

CC of the breast, fibrocystic changes, proliferative breast disease and

CC non-carcinoma tumours.

Query Match 80.0%; Score 16.8; DB 22; Length 476;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tgaagaaggaagctaaagt 21  
||||||| |  
DB 159 tgaagaagctagctaaagt 178

## RESULT 12

ID AAX40071 standard; DNA; 2569 BP.

AC AAX40071;

DT 02-JUL-1999 (first entry)

DE Colon cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX prostate cancer; ss.

OS Homo sapiens.

PN WO904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

PA (LUDWIG-) LUDWIG INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfreundschuh M, Sahlin U, Scanlan MJ, Stockert E;

PI Tureci O;

DR WPI; 1999-132448/11.

PT New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

PS Claim 67; Page 677-678; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically

CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the

CC interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in

CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides

CC which are cancer associated antigen precursors expressed in human breast

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC lung cancer.

XX Sequence 2569 BP; 865 A; 418 C; 564 G; 712 T; 10 other;

Query Match 80.0%; Score 16.8; DB 20; Length 2569;

Best Local Similarity 90.0%; Pred. No. 2.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 tgaagaagaaagctaaagt 21  
Db 1242 tgaagaagaaagctaaagt 1261

RESULT 13  
AA229782/C  
ID AA229782 standard; DNA: 222 BP.  
XX  
AC AA229782;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE Putative PMG-1 from tomato mutant UQ14.  
XX  
KW Tomato mutant UQ14; sucrose synthase gene;  
KM PMGS; phenotype modifying genetic sequence; plant pathogen resistance;  
KM senescence timing; starch metabolism; ds.  
XX  
OS Lycopersicon esculentum.  
XX  
PN WO9963068-A1.  
XX  
PD 09-DEC-1999.  
XX  
PE 04-JUN-1999; 99WO-AU00434.  
XX  
PR 04-JUN-1998; 98AU-0003901.  
PR 04-JUN-1998; 98AU-0003903.  
PR 25-SEP-1998; 98AU-0006169.  
PR 25-SEP-1998; 98AU-0006174.  
XX  
PA (UYOU ) UNIV QUEBENS LAND.  
XX  
PI Carrot 1 BJ;  
XX  
DR WPI: 2000-116368/10.  
XX  
PT New polynucleotides that increase gene expression in plants used to  
PT produce transgenic plants with resistance to plant pathogens -  
XX  
PS Claim 20; Page 63; 93pp; English.  
XX  
CC The present sequence is a phenotype modifying genetic sequence upstream  
CC of DS insertion. This shows homology to 3' untranslated region of potato  
CC sucrose synthase gene. This can be used to produce transgenic plants  
CC having altered phenotypic traits, such as resistance to plant pathogens,  
CC senescence timing, starch metabolism, cell growth, expansion and/or  
CC division, and the shape of cells, tissues or organs.  
XX  
SQ Sequence 222 BP; 63 A; 28 C; 41 G; 66 T; 24 other;

Query Match 78.1%; Score 16.4; DB 21; Length 222;  
Best Local Similarity 94.4%; Pred. No. 3.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 aaaaggaagctaaagt 21  
Db 152 AAAAGAAAAGCTAAAGT 135

RESULT 14  
AAH83805  
ID AAH83805 standard; cDNA: 346 BP.  
XX  
AC AAH83805;  
XX  
DT 25-SEP-2001 (first entry)  
XX

DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:1429.  
XX  
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151513-A2.  
XX  
PD 19-JUL-2001.  
XX  
PE 16-JAN-2001; 2001WO-US01575.  
XX  
PR 14-JAN-2000; 2000US-0176722.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Algate PA;  
XX  
DR WPI: 2001-425866/45.  
XX  
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer -  
XX  
PS Claim 5; Page 325; 338pp; English.  
XX  
CC AAH82377 to AAH83878 represent human ovarian tumour-associated  
CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit  
CC the development of cancer, particularly ovarian cancer. They can also  
CC be used to diagnose the onset and progression of cancer.  
XX  
SQ Sequence 346 BP; 90 A; 64 C; 66 G; 126 T; 0 other;

Query Match 78.1%; Score 16.4; DB 22; Length 346;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaa 18  
Db 25 atgaaaggaagctaa 42

RESULT 15  
AAH83806  
ID AAH83806 standard; cDNA: 346 BP.  
XX  
AC AAH83806;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:1430.  
XX  
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151513-A2.  
XX  
PD 19-JUL-2001.  
XX  
PE 16-JAN-2001; 2001WO-US01575.  
XX  
PR 14-JAN-2000; 2000US-0176722.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Algate PA;  
XX



DR WPI: 2001-425866/45.

XX  
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer .  
XX  
XX  
PS Claim 5; Page 325; 338pp; English.

XX  
CC AAH82377 to AAH83878 represent human ovarian tumour-associated  
CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit  
CC the development of cancer, particularly ovarian cancer. They can also  
CC be used to diagnose the onset and progression of cancer.  
XX

SQ Sequence 346 BP; 89 A; 64 C; 66 G; 126 T; 1 other;

Query Match 78.1%; Score 16.4; DB 22; Length 346;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 atgaaaggaagcctaaa 18  
||||||| |||||||  
Db 25 atgaaagcaagcctaaa 42

Search completed: March 20, 2002, 01:57:37  
Job time: 7018 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 00:45:13 ; Search time 2311.81 seconds  
(without alignments)  
97.612 Million cell updates/sec

Title: US-09-004-395-3

Sequence: 1 atgaagaaggaagctaaagt 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_estfun: \*  
2: em\_esthum: \*  
3: em\_estlin: \*  
4: em\_estcom: \*  
5: em\_estcpl: \*  
6: em\_estba: \*  
7: em\_estro: \*  
8: em\_estov: \*  
9: em\_hic: \*  
10: gb\_estl: \*  
11: gb\_est2: \*  
12: gb\_hic: \*  
13: gb\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_hum: \*  
16: em\_gss\_inv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_rod: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	857	13	CNSO1NND
2	19.4	92.4	863	13	CNSO1NUR
3	18.4	87.6	211	10	AA514077
4	18.4	87.6	225	10	AV276062
5	18.4	87.6	449	11	BG653289
6	18.4	87.6	457	13	BG653289
7	18.4	87.6	654	13	AO151824
8	18.4	87.6	717	10	AZ992563
9	18.4	87.6	887	10	BE537306
10	18.4	85.7	281	10	BE539217
11	18	85.7	308	10	BB326258
12	18	85.7	467	13	BB552761
					AZ936827
					2M0193D13

13	17.8	84.8	217	10	AA700025
14	17.8	84.8	296	11	N73600
15	17.8	84.8	311	10	AV219438
16	17.8	84.8	321	10	AL586792
17	17.8	84.8	324	10	BB494627
18	17.8	84.8	335	11	D63016
19	17.8	84.8	362	13	AO824834
20	17.8	84.8	367	13	AO146623
21	17.8	84.8	387	11	T87926
22	17.8	84.8	401	11	D62534
23	17.8	84.8	414	10	AA987196
24	17.8	84.8	427	13	AZ666735
25	17.8	84.8	428	11	R19941
26	17.8	84.8	442	10	A1694973
27	17.8	84.8	473	11	BG055061
28	17.8	84.8	478	10	A1953318
29	17.8	84.8	491	10	A1371083
30	17.8	84.8	506	10	A1589168
31	17.8	84.8	508	10	AM662149
32	17.8	84.8	511	10	A1478812
33	17.8	84.8	518	13	AZ479417
34	17.8	84.8	519	11	BF447906
35	17.8	84.8	579	10	BE669553
36	17.8	84.8	592	10	A1179333
37	17.8	84.8	598	10	AM077421
38	17.8	84.8	677	11	BG292236
39	17.8	84.8	711	10	AV716967
40	17.8	84.8	738	11	BG611921
41	17.8	84.8	754	11	BG283645
42	17.8	84.8	759	11	BF683582
43	17.8	84.8	872	13	AZ186410
44	17.8	84.8	991	13	CNS078RT
45	17.4	82.9	296	10	BB229099

#### ALIGNMENTS

RESULT 1  
CNSO1NND  
LOCUS  
DEFINITION  
CNSO1NND 857 bp DNA  
Anopheles gambiae GSS sp6 end of clone 24H10 of NotreDame library  
from strain PEST of Anopheles gambiae (African malaria mosquito),  
genomic survey sequence.  
AL152442  
VERSION AL152442.1 GI:7013361  
KEYWORDS  
SOURCE  
ORGANISM  
Anopheles gambiae  
African malaria mosquito.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
Culicoidae; Anopheles.  
1 (bases 1 to 857)  
Genoscope.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
2 (bases 1 to 857)  
Direct Submission  
Submitted (16-FEB-2000) BBMT, Institut Pasteur, 25, rue du Dr.  
Roux, Paris 75015, France  
COMMENT  
This clone is from an A. gambiae BAC library provided by F.H.  
Collins and sequenced by Genoscope in collaboration with the  
Laboratory of Biochem. and Biol. Molec. of Insects, Institut  
Pasteur.  
FEATURES  
source  
1. 857  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="24H10"

```

/clone.lib="Notredame1"
/note="end : SP6"
BASE COUNT      270 a      151 c      142 g      290 t      4 others
ORIGIN

Query Match      92.4%: Score 19.4; DB 13; Length 857;
Best Local Similarity 95.2%: Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21
|||||
Db 601 ATGAAAGGAAGCTAAAGT 621

RESULT 2
CONS1NDR      863 bp      DNA      GSS      14-JUN-2001
LOCUS      Anopheles gambiae GSS SP6 end of clone 24005 of Notredame1 library
DEFINITION      from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION      AL152708
VERSION      AL152708.1 GI:7013627
KEYWORDS      African malaria mosquito.
SOURCE      Anopheles gambiae
ORGANISM      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anopheles.
REFERENCE      1 (bases 1 to 863)
AUTHORS      Genoscope.
TITLE      Direct Submision
JOURNAL      Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : segrefigenoscope.cns.fr
Web : www.genoscope.cns.fr)
2 (bases 1 to 863)
Roch,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Direct Submision
Submitted (16-FEB-2000) BMWI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source      1..863
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="24005"
/clone.lib="Notredame1"
/note="end : SP6"
BASE COUNT      275 a      153 c      142 g      291 t      2 others
ORIGIN

Query Match      92.4%: Score 19.4; DB 13; Length 863;
Best Local Similarity 95.2%: Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21
|||||
Db 578 ATGAAAGGAAGCTAAAGT 598

RESULT 3
AA514077      211 bp      mRNA      EST      10-JUL-1997
LOCUS      MC2S032.RCS S. mansoni cercarial lambda zap Schistosoma mansoni
DEFINITION      cDNA clone AS-32 5', mRNA sequence.
ACCESSION      AA514077
VERSION      AA514077.1 GI:2252933
KEYWORDS      EST.

```

```

SOURCE      Schistosoma mansoni.
ORGANISM      Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
Rhabdiphora; Eulictophora; Revertospermatia; Medifusta;
Neodermata; Trematoda; Digenea; Strigeidae; Schistosomatidae;
Schistosomatidae; Schistosoma.
REFERENCE      1 (bases 1 to 211)
AUTHORS      Karim,A.M.
TITLE      Schistosoma Gene Identification
JOURNAL      Unpublished (1997)
COMMENT      Contact: Karim,A.M.
Dr.Amr Karim, Department of Biochemistry
Faculty of Science, Ain Shams University
Box 11456, El-Khalifa El-Maamoun St., Abbasia, Cairo, Egypt
Tel: (202)2853561
Fax: (202)2853561
Email: Hadyeasnet.shams.eun.eg
Seq primer: SK
High quality sequence stop: 211.
FEATURES
source      1..211
Location/Qualifiers
/organism="Schistosoma mansoni"
/db_xref="taxon:6183"
/clone="AS-32"
/clone.lib="S. mansoni cercarial lambda zap"
/sex="Mixed"
/dev_stage="cercarial"
/note="library provided by Dr. M. Saber TBRI,Cairo,Egypt."
BASE COUNT      61 a      36 c      51 g      63 t
ORIGIN

Query Match      87.6%: Score 18.4; DB 10; Length 211;
Best Local Similarity 95.0%: Pred. No. 3.8e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgaaggaagctaaagt 21
|||||
Db 153 TGAAGGAAGCTTAAAGT 172

RESULT 4
AV276062      225 bp      mRNA      EST      05-NOV-1999
LOCUS      AV276062/c
DEFINITION      Musculus cDNA clone 493243ZH20 3', mRNA sequence.
ACCESSION      AV276062
VERSION      AV276062.1 GI:6264099
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 225)
Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Suganara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateo,M., Tomioka,N., Tsunoda,Y.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Kono,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Science Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

```



JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3080 row: P column: 21  
Class: BAC ends  
High quality sequence stop: 457.  
Location/Qualifiers

FEATURES  
source

1. .457  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate:3080 Col-21 Row-P"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC1; BAC clones in E-Coli DH10B"

BASE COUNT 135 a 92 c 72 g 158 t

ORIGIN

Query Match 87.6%; Score 18.4; DB 13; Length 457;  
Best Local Similarity 95.0%; Pred. No. 3.1e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atgaagaagaaagctaaag 20  
|||||  
Db 163 ATGAACAGCAAGCTAAAG 144

RESULT 7  
AZ992563 654 bp DNA GSS 27-APR-2001  
LOCUS 2M0277G14F Mouse 10kb plasmid UNGC2M library Mus musculus genomic  
DEFINITION clone UNGC2M0277G14 F, DNA sequence.  
ACCESSION AZ992563  
VERSION AZ992563.1 GI:13863790  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 654)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0277 row: G column: 14  
Seq Primer: CGTGTGAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 654.  
Location/Qualifiers

FEATURES  
source

1. .654  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

/clone="UNG2M0277G14"  
/clone\_lib="Mouse 10kb plasmid UNGC2M library"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1473211419b1AF129072.1) a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 215 a 113 c 138 g 188 t

ORIGIN

Query Match 87.6%; Score 18.4; DB 13; Length 654;  
Best Local Similarity 95.0%; Pred. No. 2.8e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atgaagaagaaagctaaag 20  
|||||  
Db 557 ATGAACAGCAAGCTAAAG 576

RESULT 8  
BE537306 717 bp mRNA EST 09-AUG-2000  
LOCUS 601065480P1 NIH\_MGC\_10 Homo sapiens cDNA IMAGE:3452011 5',  
DEFINITION mRNA sequence.  
ACCESSION BE537306  
VERSION BE537306.1 GI:9765951  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 717)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM8432 row: O column: 20  
High quality sequence stop: 678.  
Location/Qualifiers

FEATURES  
source

1. .717  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3452011"  
/clone\_lib="NIH\_MGC\_10"  
/cell\_line="MGC36"  
/lab\_host="DH10B"  
/note="Organ: cervix; Vector: PCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.



BASE COUNT	57 a	67 c	40 g	116 t	1 others
ORIGIN					
Query Match		85.7%	Score 18;	DB 10;	Length 281;
Best Local Similarity		100.0%;	Pred. No. 4.8e+03;		
Matches 18;	Conservative	0;	Mismatches	0;	Indels 0;
					Gaps 0;
3 gaagaaggaagcttaagaag 20					
170 GAAAGGAAAGCTTAAG 153					
RESULT 11					
LOCUS BB552761/c					
DEFINITION BB552761 RIKEN full-length enriched, 2 days pregnant adult female					
ACCESSION BB552761					
VERSION BB552761.1					
KEYWORDS GI:9639127					
SOURCE EST.					
ORGANISM house mouse.					
Mus musculus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.					
1 (bases 1 to 308)					
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci					
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,					
Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,					
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,					
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,					
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,					
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata					
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y.,					
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya					
T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamataka, I.,					
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino					
M., Muramatsu, M. and Hayashizaki, Y.					
RIKEN Mouse ESTs (Konno, H., et al.)					
Unpublished (2000)					
Contact: Yoshihide Hayashizaki					
Laboratory for Genome Exploration Research Group, RIKEN Genomic					
Sciences Center(GSC), Yokohama Institute					
The Institute of Physical and Chemical Research (RIKEN)					
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan					
Tel: 81-45-503-9222					
Fax: 81-45-503-9216					
Email: genome-res@gsr.riken.go.jp,					
URL: http://genome.gsc.riken.go.jp/					
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasakita					
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.					
Thermosensitization and thermoinactivation of thermolabile enzymes by					
trichostatin and its application for the synthesis of full length					
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)					
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,					
Tomaru, Y., Carninci, P., Shibata, K., Ozawa, T., Muramatsu, M., Okazaki					
Y. and Hayashizaki, Y.					
Automated filtration-based high-throughput plasmid preparation					
system. Genome Res. 9 (5), 463-470 (1999)					
Carninci, P. and Hayashizaki, Y.					
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,					
19-44 (1999)					
Please visit our web site ( <a href="http://genome.ritc.riken.go.jp">http://genome.ritc.riken.go.jp</a> ) for					
further details.					
Location/Qualifiers					
1. 308					
/organism="Mus musculus"					
/db_xref="taxon:10090"					
/clone="E330001H21"					
/clone_lib="RIKEN full-length enriched, 2 days pregnant					
adult female ovary"					
/sex="female"					
/library_type="ovary"					

```

/def_stage="2 days pregnant adult"
/lab_host="DH10B"
/note="Site_1: Sall; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'] GAGCAGAGAGACGGCCGCGACATCGAGCTTTTCTTTTCTTTTAA 3'}, cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'] GAGCAGAGAGATCTCGCTTAAATTAATTAATGCCCCCCCCCCC 3'}. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      70 a      75 c      35 g      128 t
ORIGIN
Query Match      85.7%; Score 18; DB 10; Length 308;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      3 gaaaggaagactaaag 20
      |||||||
Db      138 GAAAAGGAAAGCTTAAG 121
RESULT 12
A2936827      467 bp      DNA      GSS      26-APR-2001
LOCUS      2M0193D13R Mouse 10kb plasmid UNGC2M library Mus musculus genomic
DEFINITION      clone UNGC2M0193D13 R, DNA sequence.
ACCESSION      A2936827
VERSION
KEYWORDS
SOURCE
ORGANISM
      house mouse.
      Mus musculus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      1 (bases 1 to 467)
REFERENCE
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,
      Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly
      ,M., Rose,M., Rose,R., Stokes,R., Stokes,A., von Niederhausen,A.
      and Wright,D., Weiss,R.
      Mouse whole genome scaffolding with paired end reads from 10kb
      plasmid inserts
      Unpublished (2000)
JOURNAL
COMMENT      Contact: Robert B. Weiss
      University of Utah Genome Center
      University of Utah
      Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
      84112 USA
      Tel: 801 585 5606
      Fax: 801 585 7177
      Email: ddunn@genetics.utah.edu
      Insert Length: 10000 Std Error: 0.00
      Plate: 0193 row: D column: 13
      Seq primer: CACACAGAGAAACACCTATGAC
      Class: plasmid ends
      High quality sequence stop: 467.
      Location/Qualifiers
      1..467
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UNG2M0193D13"
      /clone_1lb="Mouse 10kb plasmid UNGC2M library"
      /sex="Female"
      /lab_host="E. coli strain XL10-Gold, Ti-resistant, F-"
      /note="Vector: pMD42nv; Purified genomic DNA from M.

```



musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 92 a 111 c 146 g 115 t 3 others  
ORIGIN

Query Match 85.7%; Score 18; DB 13; Length 467;  
Best Local Similarity 100.0%; Pred. No. 4.2e+03;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aaaaagaaagctaaagt 21  
|||||  
Db 34 AAAAGGAAAGCTAAAGT 51

RESULT 13  
AA700025 217 bp mRNA EST 19-DEC-1997  
LOCUS z169d12.s1 Soares\_fetal\_liver\_spleen\_INFLS\_S1 Homo sapiens CDNA  
DEFINITION clone IMAGE:436055 3', mRNA sequence.  
ACCESSION AA700025  
VERSION AA700025.1 GI:2702988  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 217)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Getzel, G., Jost, S.,  
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](http://info@image.llnl.gov)) for further information.  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 107.

## FEATURES

Source  
1..217  
/organism="Homo sapiens"  
/db\_xref="GDB:1335824"  
/db\_xref="taxon:9606"  
/clone="IMAGE:436055"  
/clone\_lib="Soares\_fetal\_liver\_spleen\_INFLS\_S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)  
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;  
This is a subtracted version of the original Soares fetal

liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I - oligo(dT) primer [5',  
AACTGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified p773 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Felina Bonaldo."

BASE COUNT 72 a 47 c 54 g 44 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 217;  
Best Local Similarity 90.5%; Pred. No. 6.1e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgaagaaagctaaagt 21  
|||||  
Db 50 ATGTAAAGAAAGCTAAAT 70

RESULT 14  
N73600 296 bp mRNA EST 19-MAR-1996  
LOCUS za51d06.s1 Soares\_fetal\_liver\_spleen\_INFLS\_Homo sapiens CDNA clone  
DEFINITION IMAGE:296075 3' similar to contains Alu repetitive element., mRNA  
sequence.  
ACCESSION N73600  
VERSION N73600.1 GI:1230885  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 296)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,  
R., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Treviski, E., Waterston,  
R., Williamson, A., Wohldmann, P. and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](http://info@image.llnl.gov)) for further information.  
Seq primer: m13 -40 forward  
High quality sequence stop: 212.

## FEATURES

Source  
1..296  
/organism="Homo sapiens"  
/db\_xref="GDB:124096"  
/db\_xref="taxon:9606"  
/clone="IMAGE:296075"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)  
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5', AACTGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified p773 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Felina Bonaldo."

BASE COUNT 95 a 64 c 68 g 66 t 3 others  
ORIGIN







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 01:30:37 ; Search time 116.86 Seconds  
(without alignments)  
40.699 Million cell updates/sec

Title: US-09-004-395-3

Perfect score: 21

Sequence: 1 atgaaagaagactaaagt 21

Scoring table:

IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:\*

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.2	77.1	594	1	US-08-404-732A-3
C 2	16.2	77.1	1420	1	US-08-080-255-3
C 3	16.2	77.1	1420	3	US-08-465-713-3
C 4	16.2	77.1	1420	5	PCT-US93-05857-3
C 5	16.2	77.1	1853	1	US-08-404-732A-6
C 6	16.2	77.1	2108	4	US-09-032-742-6
C 7	16.2	77.1	2120	1	US-08-404-732A-4
C 8	16.2	77.1	2224	1	US-08-404-732A-8
C 9	16.2	77.1	2325	2	US-08-966-388-5
C 10	16.2	77.1	2325	3	US-09-188-403-5
C 11	16.2	77.1	2325	3	US-09-188-403-5
C 12	16.2	77.1	2325	3	US-09-188-403-5
C 13	16.2	77.1	2325	4	US-09-281-259-5
C 14	16.2	77.1	3956	2	US-08-966-388-9
C 15	16.2	77.1	3956	3	US-09-188-403-9
C 16	16.2	77.1	3956	3	US-09-188-404-9
C 17	16.2	77.1	3956	4	US-09-281-259-9
C 18	16.2	77.1	4201	1	US-08-080-255-4
C 19	16.2	77.1	4201	1	US-08-465-713-4
C 20	16.2	77.1	4201	5	PCT-US93-05857-4
C 21	16.2	77.1	11907	4	US-08-061-376-4
C 22	16.2	77.1	14255	1	US-08-320-559-1
C 23	16.2	77.1	14255	1	US-08-327-392-1
C 24	16.2	77.1	14255	1	US-08-306-691B-55
C 25	16.2	77.1	14255	3	US-08-545-860D-1
C 26	16.2	77.1	14255	5	PCT-US94-04496-1
C 27	15.8	75.2	414	4	US-09-330-330-3
					Sequence 33, Appl

28	15.8	75.2	1012	4	US-09-446-504-32	Sequence 32, Appl
29	15.8	75.2	1554	2	US-08-031-538-8	Sequence 8, Appl
30	15.8	75.2	1730	1	US-07-817-920-1	Sequence 1, Appl
31	15.8	75.2	1730	1	US-08-117-006-1	Sequence 1, Appl
32	15.8	75.2	1730	1	US-08-216-594-1	Sequence 1, Appl
33	15.8	75.2	1730	5	PCT-US93-00149-1	Sequence 1, Appl
34	15.8	75.2	2781	3	US-08-749-522-4	Sequence 4, Appl
35	15.4	73.3	4431	4	US-09-532-803-8	Sequence 8, Appl
36	15.4	73.3	12145	3	US-08-968-563-19	Sequence 19, Appl
37	15.4	73.3	12145	3	US-08-969-683A-19	Sequence 19, Appl
38	15.2	72.4	913	1	US-08-109-391A-1	Sequence 1, Appl
39	15.2	72.4	913	1	US-08-459-019A-1	Sequence 1, Appl
40	15.2	72.4	913	2	US-08-460-428A-1	Sequence 1, Appl
41	15.2	72.4	913	3	US-08-458-860A-1	Sequence 1, Appl
42	15.2	72.4	1972	1	US-08-463-048-1	Sequence 1, Appl
43	15.2	72.4	1972	1	US-08-463-229-1	Sequence 1, Appl
44	15.2	72.4	1972	2	US-08-302-891-1	Sequence 1, Appl
45	15.2	72.4	3763	1	US-07-792-865D-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-404-732A-3/C  
Sequence 3, Application US/08404732A  
Patent No. 5661017  
GENERAL INFORMATION:  
APPLICANT: Dunahay, Terry G.  
APPLICANT: Roesler, Paul  
APPLICANT: Jarvis, Eric  
TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE, MATERIALS  
TITLED OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: National Renewable Energy Laboratory  
ADDRESS: 1617 Cole Boulevard  
CITY: Golden  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80401-3393  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(DOS)text (\*.\*)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/404,732A  
FILING DATE: 15-030-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Connor, Edna M.  
REGISTRATION NUMBER: 29252  
REFERENCE/DOCKET NUMBER: 94-31  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/275-4400  
TELEFAX: 303/275-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 594 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 1..594  
US-08-404-732A-3

Query Match 77.1%; Score 16.2; DB 1; Length 594;  
Best Local Similarity 85.7%; Pred. No. 95;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
||||| ||| |||  
DB 100 ATGAGAAGCAATCTAAAGT 80

## RESULT 2

US-08-080-255-3/c  
; Sequence 3, Application US/08080255  
; Patent No. 5487970

## GENERAL INFORMATION:

APPLICANT: Rowley, Janet D.  
APPLICANT: Diaz, Manuel O.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/080.255  
FILING DATE: 19930617  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: ARCD:072/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 320-7200  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1420 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-080-255-3

Query Match 77.1%; Score 16.2; DB 1; Length 1420;  
Best Local Similarity 85.7%; Pred. No. 98;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
||||| ||| |||

DB 70 ATGAGAAGCAATCTAAAGT 50

## RESULT 3

US-08-465-713-3/c  
; Sequence 3, Application US/08465713  
; Patent No. 6121419

## GENERAL INFORMATION:

APPLICANT: Rowley, Janet D.  
APPLICANT: Diaz, Manuel O.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433

CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465.713  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/080.255  
FILING DATE: 17 JUNE 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: ARCD:072/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 320-7200  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1420 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-465-713-3

Query Match 77.1%; Score 16.2; DB 3; Length 1420;  
Best Local Similarity 85.7%; Pred. No. 98;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
||||| ||| |||

DB 70 ATGAGAAGCAATCTAAAGT 50

## RESULT 4

PCT-US93-05857-3/c  
; Sequence 3, Application PC/TUS9305857  
; GENERAL INFORMATION:

## APPLICANT: Board of Regents

APPLICANT: The University of Texas System  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING  
TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/05857  
FILING DATE: 19930617  
CLASSIFICATION:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/900,689  
FILING DATE: 17/06/92  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165

```

; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1420 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-05857-3

Query Match      77.1% Score 16.2; DB 5; Length 1420;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21
    ||| ||||| |||||
Db 70 ATGAGAGCAAGAAAGTAAAGT 50

RESULT 5
US-08-404-732A-6/c
; Sequence 6, Application US/08404732A
; Patent No. 5661017
; GENERAL INFORMATION:
; APPLICANT: Dunahay, Terry G.
; APPLICANT: Roesler, Paul
; APPLICANT: Jarvis, Eric
; TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE MATERIALS
; TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,732A
; FILING DATE: 15-03-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna M.
; REGISTRATION NUMBER: 29252
; REFERENCE/DOCKET NUMBER: 94-31
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/275-4404
; TELEFAX: 303/275-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1853 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: 1. 5'UTR
; LOCATION: 1..445
; OTHER INFORMATION: /label= Accase
; NAME/KEY: CDS
; LOCATION: 446..1240
; OTHER INFORMATION: /label= NPRTII
; NAME/KEY: Linker Sequence
```

```

; LOCATION: 1241..1243
; NAME/KEY: Accase 3' coding region (in non-translatable reading
; frame)
; LOCATION: 1244..1259
; NAME/KEY: 3'UTR
; LOCATION: 1260..1853
; OTHER INFORMATION: /label= Accase
; US-08-404-732A-6

Query Match      77.1% Score 16.2; DB 1; Length 1853;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21
    ||||| || |||||
Db 1359 ATGAAAGCAATCTAAAGT 1339

RESULT 6
US-09-032-742-6/c
; Sequence 6, Application US/09032742
; Patent No. 6255089
; GENERAL INFORMATION:
; APPLICANT: Teitler, Milt
; APPLICANT: Herrick-Davis, Katharine
; APPLICANT: Egan, Christina C.
; TITLE OF INVENTION: Constitutively Activated Serotonin
; TITLE OF INVENTION: Receptors
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Laurence Weinberger
; STREET: 882 S. Malack Street, Suite 103
; CITY: P.O. Box 1663
; STATE: PA
; COUNTRY: USA
; ZIP: 19380-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,742
; FILING DATE: 27-FEB-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinberger, Laurence
; REGISTRATION NUMBER: 27,965
; REFERENCE/DOCKET NUMBER: 3086-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 431-1703
; TELEFAX: (610) 431-4181
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2108 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-032-742-6

Query Match      77.1% Score 16.2; DB 4; Length 2108;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21
    ||||| || |||||
Db 1575 ATGAAAGCAATCTAAAGT 1555
```

```
RESULT 7
US-08-404-732A-4/c
; Sequence 4, Application US/08404732A
; Patient No. 5661017
; GENERAL INFORMATION:
; APPLICANT: Dunahay, Terry G.
; APPLICANT: Roesler, Paul
; APPLICANT: Jarvis, Eric
; TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE MATERIALS
; TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(DOS)text (*.*)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,732A
; FILING DATE: 15-030-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna M.
; REGISTRATION NUMBER: 29252
; REFERENCE/DOCKET NUMBER: 94-31
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/275-4400
; TELEFAX: 303/275-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2120 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; FEATURE: 1.
; NAME/KEY: 5'UTR
; LOCATION: 1..445
; OTHER INFORMATION: /label= Accase
; NAME/KEY: CDS
; LOCATION: 446..1240
; OTHER INFORMATION: /label= NP111
; NAME/KEY: Linker Sequence
; LOCATION: 1241..1248
; NAME/KEY: ACCase 3' coding region (in non-translatable reading
; NAME/KEY: frame)
; LOCATION: 1249..1526
; NAME/KEY: 3'UTR
; LOCATION: 1527..2120
; OTHER INFORMATION: /label= Accase
; US-08-404-732A-4

Query Match 77.1%; Score 16.2; DB 1; Length 2120;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 atgaaaggaagctaaagt 21
||||| ||| ||| |||
DB 1626 ATGAAAGCAATCCTAAAGT 1606

RESULT 8
US-08-404-732A-8/c
; Sequence 8, Application US/08404732A
; Patient No. 5661017
```

```
; GENERAL INFORMATION:
; APPLICANT: Dunahay, Terry G.
; APPLICANT: Roesler, Paul
; APPLICANT: Jarvis, Eric
; TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE MATERIALS
; TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: National Renewable Energy Laboratory
; STREET: 1617 Cole Boulevard
; CITY: Golden
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(DOS)text (*.*)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,732A
; FILING DATE: 15-030-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Edna M.
; REGISTRATION NUMBER: 29252
; REFERENCE/DOCKET NUMBER: 94-31
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/275-4400
; TELEFAX: 303/275-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2224 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; FEATURE: 1.
; NAME/KEY: 5'UTR
; LOCATION: 1..816
; OTHER INFORMATION: /label= Accase
; NAME/KEY: CDS
; LOCATION: 817..1611
; OTHER INFORMATION: /label= NP111
; NAME/KEY: Linker Sequence
; LOCATION: 1612..1614
; NAME/KEY: ACCase 3' coding region (in non-translatable reading
; NAME/KEY: frame)
; LOCATION: 1615..1630
; NAME/KEY: 3'UTR
; LOCATION: 1631..2224
; OTHER INFORMATION: /label= Accase
; US-08-404-732A-8

Query Match 77.1%; Score 16.2; DB 1; Length 2224;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 atgaaaggaagctaaagt 21
||||| ||| ||| |||
DB 1730 ATGAAAGCAATCCTAAAGT 1710

RESULT 9
US-08-966-388-5
; Sequence 5, Application US/08966388
; Patient No. 5965412
; GENERAL INFORMATION:
; APPLICANT: TOMOYUKI NISHIMOTO
; APPLICANT: MICHIO KUBOTA
; APPLICANT: HIROTO CHAEN
```



APPLICANT: TOSHIO MIYAKE  
TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/966,388  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 311,235/1996  
FILING DATE: 8-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-966-388-5

Query Match 77.1%; Score 16.2; DB 2; Length 2325;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgaaaaggaagctaaagt 21  
1 ||||| ||||| |||||  
Db 2274 AGCAAAAGCAATATATAAGT 2294

RESULT 10  
US-09-188-403-5  
Sequence 5, Application US/09188403  
Patent No. 6066477  
GENERAL INFORMATION:  
APPLICANT: TOMOYUKI NISHIMOTO  
APPLICANT: MICHIO KUBOTA  
APPLICANT: HIROO CHAEN  
APPLICANT: TOSHIO MIYAKE  
TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/188,403  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/966,388  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-188-403-5

Query Match 77.1%; Score 16.2; DB 3; Length 2325;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgaaaaggaagctaaagt 21  
1 ||||| ||||| |||||  
Db 2274 AGCAAAAGCAATATATAAGT 2294

RESULT 11  
US-09-188-404-5  
Sequence 5, Application US/09188404  
Patent No. 6140487  
GENERAL INFORMATION:  
APPLICANT: TOMOYUKI NISHIMOTO  
APPLICANT: MICHIO KUBOTA  
APPLICANT: HIROO CHAEN  
APPLICANT: TOSHIO MIYAKE  
TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/188,404  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/966,388  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-188-404-5

Query Match 77.1%; Score 16.2; DB 3; Length 2325;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaagaagaaagctaaagt 21  
| |||||  
Db 2274 AGGAAAGGAATATATAAGT 2294

RESULT 12  
US-09-281-259-5  
; Sequence 5, Application US/09281259  
; Patent No. 6204377  
GENERAL INFORMATION:  
APPLICANT: TOMOYUKI NISHIMOTO  
APPLICANT: MICHIO KUBOTA  
APPLICANT: HIROTO CHAEN  
APPLICANT: TOSHIO MIYAKE  
TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/281,259  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/966,388  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-281-259-5

Query Match 77.1%; Score 16.2; DB 4; Length 2325;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaagaagaaagctaaagt 21  
| |||||  
Db 2274 AGGAAAGGAATATATAAGT 2294

RESULT 13  
US-08-966-388-9  
; Sequence 9, Application US/08966388  
; Patent No. 5965412  
GENERAL INFORMATION:  
APPLICANT: TOMOYUKI NISHIMOTO  
APPLICANT: MICHIO KUBOTA  
APPLICANT: HIROTO CHAEN  
APPLICANT: TOSHIO MIYAKE  
TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/966,388  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 311,235/1996  
FILING DATE: 8-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3956 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Thermoanaerobium brockii  
STRAIN: ATCC 35047  
FEATURE:  
FEATURE: 1-100 5'-UTR  
FEATURE: E  
FEATURE: 101-2425 mat peptide  
FEATURE: S  
FEATURE: 2426-3956 3'-UTR  
FEATURE: E  
US-08-966-388-9

Query Match 77.1%; Score 16.2; DB 2; Length 3956;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Oy      1 atgaaaagyaagctaaagt 21
          | ||||| ||| |||
Db      2374 AGGAAAAGGAAATATAAAGT 2394

```

```

1 RESULT 14
2 US-09-188-403-9
3 : Sequence 9, Application US/09188403
4 : Patent No. 6064477
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: TOMOYUKI NISHIMOTO
8 : APPLICANT: Michio KUBOTA
9 : APPLICANT: Hiroto CHAEN
10 : APPLICANT: Toshio MIYAKE
11 : TITLE OF INVENTION: KOIIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES
12 : NUMBER OF SEQUENCES: 10
13 :
14 : CORRESPONDENCE ADDRESS:
15 : ADDRESSEE: BROWDY AND NEIMARK
16 : STREET: 419 Seventh Street, N.W., Suite 300
17 : CITY: Washington
18 : STATE: D.C.
19 : COUNTRY: USA
20 : ZIP: 20004
21 :
22 : COMPUTER READABLE FORM:
23 : MEDIUM TYPE: Floppy disk
24 : COMPUTER: IBM PC compatible
25 : OPERATING SYSTEM: PC-DOS/MS-DOS
26 : SOFTWARE: Patent In Release #1.0, Version #1.30
27 : CURRENT APPLICATION DATA:
28 : APPLICATION NUMBER: US/09/188,403
29 : FILING DATE:
30 : CLASSIFICATION:
31 : PRIOR APPLICATION DATA:
32 : APPLICATION NUMBER: 08/966,388
33 : FILING DATE:
34 :
35 : PRIOR APPLICATION DATA:
36 : APPLICATION NUMBER: JP 61,710/97
37 : FILING DATE: 3-MAR-1997
38 : ATTORNEY/AGENT INFORMATION:
39 : NAME: BROWDY, Roger L.
40 : REGISTRATION NUMBER: 25,618
41 : TELECOMMUNICATION INFORMATION:
42 : TELEPHONE: 202-628-5197
43 : TELEFAX: 202-737-3528
44 : TELEX: 248633
45 : INFORMATION FOR SEO ID NO: 9:
46 : SEQUENCE CHARACTERISTICS:
47 : LENGTH: 3956 base pairs
48 : TYPE: nucleic acid
49 : STRANDEDNESS: double
50 : TOPOLOGY: linear
51 : MOLECULE TYPE: Genomic DNA
52 : ORIGINAL SOURCE:
53 : ORGANISM: Thermotoga maritima
54 : STRAIN: ATCC 35047
55 :
56 : FEATURE:
57 : FEATURE: 1-100 5'-UTR
58 : FEATURE: E
59 : FEATURE: 101-2425 mat peptide
60 : FEATURE: S
61 : FEATURE: 2426-3956 3'-UTR
62 : FEATURE: E
63 : US-09-188-403-9

```

Query Match	77.1%	Score 16.2	DB 3	Length 3956
Best Local Similarity	85.7%	Pred. No. 1e+02		
Matches 18	Conservative	0	Mismatches 3	Indels 0
Gaps 0				
1 atgaaagaagactaaagt	21			

```

Db      2374  AGGAAAAGCAATATTAAGCT  2394

RESULT  15
US-09-188-404-9
: Sequence 9, Application US/09188404
: Patent No. 6140487
:
GENERAL INFORMATION:
: APPLICANT: TOMOYUKI NISHIMOTO
: APPLICANT: MICHIO KUBOTA
: APPLICANT: HIROTO CHAEN
: APPLICANT: TOSHIO MIYAKE
: TITLE OF INVENTION: KOLIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: BROWDY AND NETMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: City: Washington
: STATE: D.C.
: COUNTRY: USA
: Zip: 20004
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/188,404
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/966,388
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 61,710/97
: FILING DATE: 3-MAR-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
:
INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3956 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Thermotoga aerophilum brockii
: STRAIN: ATCC 35047
:
FEATURE:
: FEATURE: 1-100 5'-UTR
: FEATURE: E
: FEATURE: 101-2425 mat peptide
: FEATURE: S
: FEATURE: 2426-3956 3'-UTR
: FEATURE: E
:
US-09-188-404-9

```

Query Match	77.1%	Score 16.2	DB 3	length 3956
Best Local Similarity	85.7%	Pred. No. 1e+02		
Matches 18; Conservative	0;	Mismatches	3;	Indels 0;
OY	1	atgaaaggaagctaaagt	21	
Db	2374	AGGAAAGGAAATATATAAGT	2394	

Search completed: March 20, 2002, 01:30:39  
Job time: 6505 sec

---

```

0Y 1 MKRKAASLIFLFLSVLPAQOETDGLAESKRAEGEVLDAEALRDSSTRLDTNVND 60
    1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKR-----FPAILGAALFVNGSAAFEQA-----FLDPSKLVGE-GMTGLAAPTID 47
    1 : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 61 YVVSASGVLPED---WVVDLGINNSVLLTPSARLQAVYKNSVYAPVAYKSESKRAG 11
    6 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 48 YSRQAGSA-YEAEDBAAKAISLAIPSWELASSQVETENOTLSLVTAPAKODARAGG 100
    4 : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 118 DTILGVRFLPFSYS-QSSAMITMPEFKIPFY-----SGESGNOFLGKGLINIKTMK 167
    1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 ETVMWVRHIFPFSFGINSFAVLIKPPITIAVYTLGATQONWAVAGQFQDFGVFLKNVGIY 166
    1 : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 168 EIKSVSYSLGVEIDLEVLFEEEDMGNGEYVASGTLKFKGMADLWMSNPYIPINISSRIKD 227
    1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 167 SIQIITLQRNLNRLISLLEQNGDERITWGIYLFQDMSKSLQNNNPYIOTVEVNRDLQ- 222
    1 : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 228 DVNPPLASSKMRKFAFVNSHSSKRYENLTFYVKDLRVLYDKLSVDSIDSDISESV-- 284
    1 : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 226 IVPYPSRPAFLIKLGKIKHRDQSGGDIVSIRKIDKIVYDQAVVNRNSVDDEALMGI 285  
QY 285 -----FKVYETSGTESLRKLAHETEK 306  
Db 286 LKREBOYRNELAKIGNLQVLRSLKKK 314

## RESULT 2

Q9LA17 PRELIMINARY: PRT: 285 AA.

AC Q9LA17: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DR 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE FLAA-1 (FRAGMENT).  
CN FLAA-1.  
OS Spirochete str. PFRS-A.  
OC Bacteria; Spirochaetales; Spirochaetaceae.  
OX NCBI\_TaxID=120784;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=PFRS-A.  
RA Sato T., Kuramitsu H.K.;  
RT "Polymerase chain reaction for the detection of flaa-1 genes of oral  
RT Spirochetes in advanced periodontal pockets."  
RL Arch. Oral Biol. 0:0-0(2000).  
DR EMBL: AF132816; AAF64553.1; -  
FT NON-TER 1  
FT SEQUENCE 285 AA: 31864 MW: 4D7FCC4BF5483C9F CRC64;

Query Match 20.0%; Score 343; DB 2; Length 285;  
Best Local Similarity 29.4%; Pred. No. 2,7e-16;  
Matches 90; Conservative 56; Mismatches 114; Indels 46; Gaps 8;

QY 13 ISTVFAOETDGLAGSKRAEPGLVDFAEIARDPSSRLDLTNYVDYVYSGAGIYKP 72  
Db 2 LNADIMADKSGGTMNRR-----TVLDYASLAD-----TSYTD-----EQK 37  
QY 73 EDMVVDLGINNMSVLLTPSARLQAVYKNSVAPAVYKSKSKRYAGTILGVRVLPSPY-S 131  
Db 38 ALMRSSLAQAQVEVNLNSARNPVAAHARSVLEAPVSEGAQFAGRGVLPPTWDS 97  
QY 133 OSSAMIMPEFKIPFYK-----GSGN-----QPL-GKGLIDNKTKEIKY 171  
Db 98 NNNAMIKAPFVIPAYEVMAQVDDQGNVQAPTEEEKASGKGRFEDGQGVKKNVGLAKSTAV 157  
QY 172 SVYSLGVEIDLEVLFDNMGMEYAVSMGTLFKFGWADLIWNPNTIPNISRILKDDVFN 231  
Db 158 NTYGMNYPHGLVMMRDQGEVHRFYMGYLLFDQSWKELVWNNPSYISDVRSREVR-LYPV 216  
QY 232 YPLASSKMFKAFRVSKSSKSKVKNFIFVVKDLRVLYDKLSIDSDISESEVFKYETS 291  
Db 217 YASPPHVEEFGFWYTRDAHAAGDYGVEFKDKVITLYDKAVLSTVRDEADDELWGIAQR 276  
QY 292 GTESLR 297  
Db 277 EAERRR 282

## RESULT 3

Q9LA16 PRELIMINARY: PRT: 285 AA.

AC Q9LA16: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DR 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE FLAA-1 (FRAGMENT).  
CN FLAA-1.  
OS Spirochete str. PFRS-B.  
OC Bacteria; Spirochaetales; Spirochaetaceae.  
OX NCBI\_TaxID=120785;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PFRS-B.  
RA Sato T., Kuramitsu H.K.;  
RT "Polymerase chain reaction for the detection of flaa-1 genes of oral  
RT Spirochetes in advanced periodontal pockets."  
RL Arch. Oral Biol. 0:0-0(2000).  
DR EMBL: AF132817; AAF64554.1; -  
FT NON-TER 1  
FT SEQUENCE 285 AA: 31917 MW: C64874F0387DBED8 CRC64;

Query Match 19.7%; Score 339; DB 2; Length 285;  
Best Local Similarity 30.9%; Pred. No. 5e-16;  
Matches 87; Conservative 49; Mismatches 104; Indels 42; Gaps 6;

QY 38 VLDFAELARDPSSRLDLTNYVDYVYSGAGIYKPEDVYDGINNMSVLLTPSARLQAV 97  
Db 21 VLDYASLAD-----TSYTD-----EQKALMRSSLAQAQVEVNLNSARNPVA 62  
QY 98 VKNSTVAPAVYKSESKRYAGDTILGVRVLPSPY-SOSSAMIMPEFKIPFY----- 146  
Db 63 HAASRVLEAPVSEGAQSFAGRGVLPPTWDSNANAMIKAPFVIPAYEVMAQVDDQ 122  
QY 147 -----SGSGNOFLGKGLIDNKTKEIKVSYSLGVEIDLEVLFDNMGMEYA 195  
Db 123 NVQAPPEEEKASGK-GRENGYGVYKKNVGLAKSTAVNTYGMNYPHGLVMMRDQGEVHR 181  
QY 196 YSMGTLFKFGWADLIWNPNTIPNISRILKDDVFNYPPLASSKMFKAFRVSKSSKVK 255  
Db 182 YFMGYLLFDQSWKELVWNNPSYISDVRSREVR-LYPVYASPPHVEEFGFWYTRDAHAAG 240  
QY 256 NFIFYVKDLRVLYDKLSIDSDISESEVFKYETSGTESLR 297  
Db 241 DIVGYFKDKVITLYDKAVLSTVRDEADDELWGIAQRREARRK 282

## RESULT 4

Q9LA18 PRELIMINARY: PRT: 285 AA.

AC Q9LA18: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DR 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE FLAA-1 (FRAGMENT).  
CN Treponema vincentii.  
OS Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=69710;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35580;  
RA Sato T., Kuramitsu H.K.;  
RT "Polymerase chain reaction for the detection of flaa-1 genes of oral  
RT Spirochetes in advanced periodontal pockets."  
RL Arch. Oral Biol. 0:0-0(2000).  
DR EMBL: AF132815; AAF64552.1; -  
FT NON-TER 1  
FT SEQUENCE 285 AA: 31830 MW: 4270CC4FAFA83C9F CRC64;

Query Match 19.6%; Score 337; DB 2; Length 285;  
Best Local Similarity 29.1%; Pred. No. 6.9e-16;  
Matches 89; Conservative 56; Mismatches 115; Indels 46; Gaps 8;

QY 13 ISTVFAOETDGLAGSKRAEPGLVDFAEIARDPSSRLDLTNYVDYVYSGAGIYKP 72  
Db 2 LNADIMADKSGGTMNRR-----TVLDYASLAD-----TSYTD-----EQK 37  
QY 73 EDMVVDLGINNMSVLLTPSARLQAVYKNSVAPAVYKSKSKRYAGTILGVRVLPSPY-S 131

```
DB 38 ALMRSSIAVAQMEVLNNSARNPVAHAASHVIEAPVSEKASFGAEVLTGVRLVLEPTWDS 97
QY 132 QSSAMIMPEKRIPEYS-----GSSGN-----QFL-GKGLDINIKTMEIKY 171
DB 98 NAAAMKRPAYIAEYMAVDQGNQAPTEEEKASGKGRFEDGGVGNVGLASIAV 157
QY 172 SVSLGVEIDLEVLFEEDMNGMEYASMGTLKFKGMADLINSNPYIPNISRIIKDDVPN 231
DB 158 NNYGMNYPHGLYVMBDQDEVRHRTFMGYLLFDSKVELVNNPSYISDVRSREVR-LYPV 216
QY 232 YPLASSKMRKAFRVSSKSHSKVKNFIYVKDLRYLTKLSVSDSDISESVFKYETS 291
DB 217 YPASTHYVVEGIMVTRDAHAAGDYGVYKDYKIIYDKAVLSTVRDFADEDLMGIDARR 276
QY 292 GTESLR 297
DB 277 EAERKR 282

RESULT 5
006691 PRELIMINARY: PRT: 243 AA.
AC 006691;
DB 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
FLAA HOMOLOG-1.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RA Porcella S.F., Radolf J.D., Norgard M.V.;
RL Submitt. (Apr-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U97363; AAB63367.1; -.
SQ SEQUENCE 243 AA; 27299 MW; 89DA106A2710B41 CRC64;

Query Match 7.6%; Score 131.5; DB 2; Length 243;
Best Local Similarity 22.5%; Pred. No. 0.11;
Matches 56; Conservative 33; Mismatches 109; Indels 51; Gaps 8;

QY 48 PSTRLDITNYVDYV---YSGASGIKPEDMVVDLGINM---SVLLTPSARLQAYKN 100
DB 21 PLAQSKSVNYQAYFFIDDFDGAS-----EDQGL-AMRAAGSKRTTGFPILKYFEG 70
QY 101 SVAPAVVSESKRYAGDTILGVRLVPPSYSSQSSAMIMP-----PKRIPYSGESGNOFL 155
DB 71 MPQAVRMAGSMQCKDKREARFIEGCKFNROGNMMLDILPTKGSDEYIPLRGVYSG---- 126
QY 156 GKGLDINIKTMEIKYSVSLGVEIDLEVLFEEDMNGMEYASMGTLKFKGMADLINSNP 215
DB 127 -----FDVWVMGAGYQYSLVALVRCOTGRVHTLLIGNLDFQGMKNLSVSVPT 173
QY 216 YIPNISRIIKDDVPNYPPLASSKMRKAFRVSSKSHSKVKNFIYVKDLRYLTKLSVSI 275
DB 174 HIQOTS-----RYLSAQLHSFVGFRIRTSPSERVDYVFDFQKAL--ANMHI 221
QY 276 DSDIDSESV 284
DB 222 DFTIDGHEL 230

RESULT 6
P96127 PRELIMINARY: PRT: 242 AA.
AC P96127;
DB 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
OS OUTER MEMBRANE PROTEIN.
GN TROMP2.
DB 276 DSDIDSESV 284
DB 222 DFTIDGHEL 230
OS Treponema pallidum.
```

```
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=97175551; PubMed=9023206;
RA Champion C.I., Blanco D.R., Exner M.M., Erdjument-Bromage H.,
Hancock R.E., Tempst P., Miller J.N., Lovett M.A.;
RT "Sequence analysis and recombinant expression of a 28-kilodalton
Treponema pallidum subsp. pallidum rate outer membrane protein
(tromp2).";
RL J. Bacteriol. 179:1230-1238(1997).
DR EMBL: U65743; AAB47846.1; -.
SQ SEQUENCE 242 AA; 27234 MW; 86C71D3136862AF8 CRC64;
```

```
Query Match 7.3%; Score 125.5; DB 2; Length 242;
Best Local Similarity 22.4%; Pred. No. 0.28;
Matches 55; Conservative 31; Mismatches 104; Indels 55; Gaps 8;

QY 48 PSTRLDITNYVDYV---YSGASGIKPEDMVVDLGINM---SVLLTPSARLQAYKN 100
DB 21 PLAQSKSVNYQAYFFIDDFDGAS-----EDQGL-AMRAAGSKRTTGFPILKYFEG 70
QY 101 SVAPAVVSESKRYAGDTILGVRLVPPSYSSQSSAMIMP-----PKRIPYSGESGNOFL 155
DB 71 MPQAVRMAGSMQCKDKREARFIEGCKFNROGNMMLDILPTKGSDEYIPLRGVYSG---- 126
QY 156 GKGLDINIKTMEIKYSVSLGVEIDLEVLFEEDMNGMEYASMGTLKFKGMADLINSNP 215
DB 127 -----FDVWVMGAGYQYSLVALVRCOTGRVHTLLIGNLDFQGMKNLSVSVPT 173
QY 216 YIPNISRIIKDDVPNYPPLASSKMRKAFRVSSKSHSKVKNFIYVKDLRYLTKLSVSI 275
DB 174 HIQOTS-----RYLSAQLHSFVGFRIRTSPSERVDYVFDFQKAL 217
QY 276 DSDID 280
DB 218 NMHD 222
```

```
RESULT 7
083669 PRELIMINARY: PRT: 242 AA.
AC 083669;
DB 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
OS OUTER MEMBRANE PROTEIN, PUTATIVE.
GN TP0663.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=9832770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
McDonald L., Artlisch P., Bowman C., Cotton M.D., Fujii C., Garland S.,
Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete.";
RL Science 281:375-388(1998).
DR EMBL: AE001240; AAC65636.1; -.
DR TIGR: TP0663; -.
KW Complete proteome.
SQ SEQUENCE 242 AA; 27257 MW; 86C71D313FEFC06F CRC64;
```





```
DB 280 LSTAGTTEIDSLFEGIDFYTTTTRARPEELNMDLFRKCMPEVEKCLDADKDKSSVMD 339
QY 287 VYETSGTESLRKLAHEFTKRVKLEKISIAEGSFONFEKIESKEEPS 337
DB 340 VVVVGG-----TRIPKVOQLVO-----DEFNCKELCKSINPDEA 374

RESULT 10
Q9ZIQ2 PRELIMINARY; PRT; 1805 AA.
AC Q9ZIQ2;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE RTX PROTEIN (APXIV).
GN APXIVA OR APXIV.V1.
OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_Taxid=715;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4074;
RA Kuhn R., Anderson T.J., Segers R.A.P.M., MacInnes J.I., Nicolet J.,
RT Frey J.;
RT "Characterization of a new RTX determinant (apxiv) of Actinobacillus
RT pleuropneumoniae.";
RT Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=4074 (SEROTYPE 1 REFERENCE STRAIN);
RA Frey J., Segers R.P.;
RT "Live attenuated bacteria of the species Actinobacillus
RT pleuropneumoniae.";
RT Patent number EP0875574-A/5, 04-NOV-1998. AKZO NOBEL NV (NL).
DR EMBL; AF021919; AAD01698.1; -.
DR EMBL; AX002409; CAB71145.1; -.
DR InterPro: IPR001343; Hemlyns_Ca_bind.
DR InterPro: IPR003355; RTX_N.
DR Pfam; PF00353; hemolysinCbind; 5.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00330; HEMOLYSIN_CALCIUM; UNKNOWN_2.
SQ SEQUENCE 1805 AA; 202129 MW; A9076E3B351BD16 CRC64;

Query Match 6.4%; Score 110; DB 2; Length 1805;
Best Local Similarity 21.5%; Pred. No. 64;
Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;

QY 22 TQGLAEGSRARPEGLVDFAE---LARDPSSTR---LDLTNYVDVYSGASIVAPEDM 75
DB 1008 TELRNSTRKSGENALVDLSELYSFNDPTKKEGLLSRID--YAKAQSFEY---- 1061
QY 76 VVDLGINNMSVLTTSARLQAYVKNSSVA---PAVVKSESKRYADDTLIGVLEPSSYS 131
DB 1062 -----NM-----AATSNLTARLRAGVIFAESPDKD----- 1090
QY 132 QSSAMIMPFKIPFYSGEGNOFL--GKGLINDIKTKMEIKVSVSLGYEIDLEVLPEDM 189
DB 1091 EKNNNLLIGSQKNNLNSGAGDILLGEGE--NDTLKSGYADRYIFSKHGQD--IYEDPT 1147
QY 190 NCMEEVAYSNGTLKFKGMADLIMSNPYINISRIIKDDVPNPPLASSKMRFAFVSKS 249
DB 1148 NNDNRARDIDITLKF-----TDVNY--AEVKFRVNDL-----MLFGYHDT 1186
QY 250 HSSKVKNFIFYVLDKLVLDKLSVSDIDSESVFK-----VYETSGTESLRKLAHEFT 305
DB 1187 DSVTVKSFYSVD---YQDPKLEFA--DRSITRDELKAGLHLGTGQND--IKDHADM 1239
QY 306 KRVLKREKISIAEGSF 322
DB 1240 DSILEGKGNCDILRGY 1256
```

```
RESULT 11
Q9TTJ4 PRELIMINARY; PRT; 906 AA.
AC Q9TTJ4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H1.
GN ITI HCL.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Suzuki A.;
RT "The sequence of rabbit inter-alpha-trypsin inhibitor heavy chain H1
RT CDNA.";
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB035707; BAA88322.1; -.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR002035; WMFA.
DR Pfam; PF00092; vwa; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
DR PROSITE; PS0234; WMFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 906 AA; 100243 MW; 02D99DF585AB7BF2 CRC64;
```

```
Query Match 6.4%; Score 109.5; DB 6; Length 906;
Best Local Similarity 18.9%; Pred. No. 25;
Matches 78; Conservative 67; Mismatches 130; Indels 137; Gaps 19;

QY 1 MKRKASILFLLSTVLFQAETDGLAEGSKRAPEGLVDFAELARDPSSTRDLITNYVD 60
DB 281 LKMKMSLVFVI-----DISGMEGQVKQTKELLKL----- 314
QY 61 VYVSGASGIKVPED---MYVDLGINNMSVLTTPS--ARLQA---YKSNVAPAVYKSE 111
DB 315 -----GDIRPEYFDLVYFGSRVOSWRSGLVPASEANLQARDYQRFSLGAT---- 363
QY 112 SKRYAGDTLIGVVL-----PFSYSSSAMIMPFKIPFYSGEGNOFLGKGLINDIKT 165
DB 364 --NLNGCLRGIEILLNNAOGLPVAASKHAITL-----IMTDEP-----TEGVTRPQI 411
QY 166 MKEIKVSV-----VSLGYEIDLEVLFEEDMNGMEVAYSMGTLKFGKGMADLIMSN----- 213
DB 412 LKNIRSAIGRFPYLSLGFCHDIDFNFLLKSLSMEN-----NGMAQRITVEDHDAQO 462
QY 214 -PNYITNISRIIKDDVPNP---LASSKMRKAF-----RVSKSHSKYKNFT 258
DB 463 LQGFYQVANPPLVLDVDELLYPQDAVVALTQHRKQYDGSSEIIVAGRIADHKLGSPK--- 519
QY 259 FYYKDLRVLYDKLSVSDIDSESVFKYETSG---TESLRKLAHEFTKRVL----- 309
DB 520 ---ADVRRAGEGQEPQTTCLVDEEMKLLRENGHMLNHNVERLMAYLTQLLELARRVKA 576
QY 310 KLRKESIAEGSFQ-----NFVKIES-----EKPESSP 339
DB 577 KGEKANVSSEALKMSLAVQFVPLTSMTRGMADEGLEPTIDKPEDSQP 628

RESULT 12
Q9LHAB PRELIMINARY; PRT; 650 AA.
AC Q9LHAB;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 70 KDA HEAT SHOCK PROTEIN (HEAT SHOCK PROTEIN 70).
GN T2E22.11.
```

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:  
OC eurosids II: Brassicales: Brassicaceae: Arabidopsis.  
NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA:  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.:  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA:  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.:  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.  
RT TAC and BAC clones.";  
RL DNA Res. 7:217-221(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, COLUMBIA:  
RX MEDLINE=21016720; PubMed=11130713;  
RA Salanoubat M., Lemcke K., Rieger M., Ansoir W., Unsel M.,  
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermayer B.,  
RA Desjany M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,  
RA De Simone V., Choisme N., Artiguenave F., Robert C., Brottier P.,  
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Queller F.,  
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
RA Wunmbach E., Drzonex H., Erle H., Jordan N., Bangert S.,  
RA Mledjman R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,  
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordiek G.,  
RA Reichelt J., Scharf M., Schoen O., Barques M., Terol J., Climent J.,  
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,  
RA Cooke R., Lande M., Berger-Liauro C., Purnelle B., Masny D.,  
RA De Haan M., Maarse A.C., Alcaraz J.-P., Cochet A., Casacuberta E.,  
RA Monfort A., Argilou A., Flores M., Liguori R., Vitale D.,  
RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
RA Rooney T., Rizzo M., Walts A., Uterback T., Fujil C.Y., Shea T.P.,  
RA Creasy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,  
RA Pal G., Mllitscher J., Sellers P., Gill J.E., Feldblum T.V.,  
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakayama S., Nakazaki N., Shimo S., Takeuchi C., Wada T.,  
RA Nakabe A., Yamada M., Yasuda M., Tabata S.:  
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
thaliana";  
RL Nature 408:820-822(2000).  
DR EMBL, AF002055; BAB02269.1;  
DR EMBL, AC069474; AAG51030.1;  
DR InterPro: IPR001023; HSP70.  
DR Pfam: PF00012; HSP70; 1.  
DR PRINTS: PR00301; HEATSHOCK70.  
DR PROSITE: PS00297; HSP70.1; 1.  
DR PROSITE: PS00329; HSP70.2; 1.  
DR PROSITE: PS01036; HSP70.3; 1.  
KW Heat shock.  
SQ SEQUENCE 650 AA; 71101 MW; 33381E96461C5B9 CRC64;

Query Match 6.3%; Score 108.5; DB 10; Length 650;  
Best Local Similarity 19.7%; Pred. No. 18;  
Matches 69; Conservative 58; Mismatches 131; Indels 93; Gaps 14;

QY 43 ELARPSSRLDITNVYVYVYGAGIYKPEMDVLDLGNW--SVLLTPSARLQAYYKN 100  
DB 61 OYANPNTIVDPARKLRIGRS-----DPSVQADKSHMPFVYVSGGK----- 104  
QY 101 SYVAPAVV---KSESKRYAGDTILIGRVLPFSYSOSSAMIMPPFR-----IPFSGESGN 152

DB 105 ----PMTVNHKGEKQSAEISSM-VLTKMEIAERFLASPLKNAVVIYPAFENDSR 159  
QY 153 QFL-GKGLIDINKMKET-KVSVSLGEYIDLE-----VLPEDNMGMEYASMKOT 202  
DB 160 QATKADAGVIGSLNMRRIINEPTAAIAVGLDKKASVGEKNVLLFDGGGFDVSLTIE 219  
QY 203 -----FKGADLIWSPNPIPNISSRIKIDDPVNYPLASSKMKFAFRYSKS 249  
DB 220 EGIEFVKATAGDTLHGDEPDNRVNHVQEFKRRKKKDDITGNPRALRLRTACERAKRT 279  
QY 250 HSKVKNFI-----FVYKDLRVLYDKLSVSI-----DSIDISESVFK 286  
DB 280 LSSAQTTEIDSLFEGIDFTTTRAFEEELNDLFKCKMEPEVEKICIRDAKMKSSVHD 339  
QY 287 VYENSGTESLRKLKAHETFKRVALKLREKISIAEGSFQNFVERISEKPEES 337  
DB 340 VVLVIGS-----TRIPVQQLQ---DFNGKRLGCSINDEA 374

RESULT 13  
Q9HK03  
ID ID Q9HK03 PRELIMINARY; PRT; 1124 AA.  
AC Q9HK03;  
DT 01-MAR-2001 (TRMBLrel. 16, Created)  
DT 01-MAR-2001 (TRMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TRMBLrel. 16, Last annotation update)  
DE HYPOTHEICAL MEMBRANE PROTEIN.  
GN TA0543.  
OS Thermoplasma acidophilum.  
OC Archaea: Euryarchaeota: Thermoplasmatales: Thermoplasmaceae;  
OC Thermoplasma.  
NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 1728;  
RX MEDLINE=20479972; PubMed=11029001;  
RA Ruepp A., Grail W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.:  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
acidophilum";  
RL Nature 407:508-513(2000).  
DR EMBL, AL445064; CAC11683.1;  
KW Hypothetical protein: Complete proteome.  
SQ SEQUENCE 1124 AA; 126534 MW; ADCEC272CC3FC2C9 CRC64;

Query Match 6.3%; Score 107.5; DB 1; Length 1124;  
Best Local Similarity 22.0%; Pred. No. 48;  
Matches 76; Conservative 38; Mismatches 95; Indels 137; Gaps 18;

QY 25 LAEGSKRAEPGE-----LVLPFAELARDPSSTRILDITN---YVDY-----VY 63  
DB 809 LINGSTYVPSGINSILSTSGIYNITAIALSTYLRNFRITDVLSTTLIDYIRYSKVY 868  
QY 64 SGASGIVKPEMDVLDGINNSVLLTPSARLQAYYKNSVAPVYKSESKRY---AGDTI 120  
DB 869 INASGYADEFE-----IKMNSYYVYVAGDTI 895  
QY 121 LGVRVLPFSYSOSSAMIMPPKIFP--YSGESGN-----QFLGKRL-IDNINKTMEIKVVS 172  
DB 896 -----HIPEGLYSIRAEVNGYSEYTGKNSIDDSALINITFS 934  
QY 173 ----VYSLGEYIDLEVLPEDNMGMEYAYSMGTLKFKGADLIWSPNPIPNISSRIIRD 228  
DB 935 RLRFMIHTATIDPADIVG--NTRFYGONFSLVDGYNDLSISRPGI----- 980  
QY 229 VPNYPLASSKMKFAFRYSKSHSKVKNFIYVYKDLR-VLYDKLSVSDSDISESVF-- 285  
DB 981 ---YPISEDSV-----LNSNMTLEFA-----LRAILV---KITIEADVDSFTIYAG 1020  
QY 286 -KYVETSGTESLRKLKAHETFRVRLKLEKISIAEGSFQNFVERIE 330

DB 1021 NKTYSVSGNDLTIYLPYGE-----HSIIQKDGYSNIFETID 1057

RESULT 14  
ID 014789 PRELIMINARY: PRT: 947 AA.  
AC 014789:  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE BRDT.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98035876; PubMed=9367677;  
RA Jones M.H., Numata M., Shimane M.;  
RT "Identification and characterization of BRDT: A testis-specific gene  
related to the bromodomain genes RING3 and Drosophila fsh.";  
RL Genomics 45:529-534(1997).  
DR EMBL; AF019085; AAB87862.1; -.  
DR InterPro: IPR001487; Bromodomain.  
DR Pfam: PF00439; bromodomain; 2.  
DR PRINTS: PR00503; BROMODOMAIN.  
DR PROSITE: PS00633; BROMODOMAIN\_1; 2.  
DR PROSITE: PS50014; BROMODOMAIN\_2; 2.  
DR SMART: SMO0297; BROMO; 2.  
SQ SEQUENCE 947 AA; 107891 MW; EC6AE07C45863783 CRC64;

Query Match 6.2%; Score 107; DB 4; Length 947;  
Best Local Similarity 18.5%; Pred. No. 40;  
Matches 76; Conservative 80; Mismatches 120; Indels 134; Gaps 20;

QY 1 MKRKKSLIFLLSTVLFQNDGLAEGSKRAPEGLVDLFDLADPEST-RDLDTNY 59  
DB 118 MDALEKLELMOKLSQMPQEOYGVKRIKGTQONIAVSAKESPSATEKVFQOET 177  
QY 60 DYVYSGASGIVKPEDMVVDLGINMSVLLTPSARLQAYVK--NSVPAVAVKSESKRY 115  
DB 178 PSVFPRTS--ISPLANYQASVNSS--QTAQVTKGVRKADTTTPATSAKASSE-- 230  
QY 116 AGDTLIGVNLFPSSQSSAMTMPK-----IPFYGESGNOFLGGLDNIKTME 168  
DB 231 -----FSPTFEKS-VALPIKENMPKVLTP-----DSQOQ--NVETVATVEQ 272  
QY 169 IK-----VSYSLG-----YEL-----DLEVLFDMMNGMEY- 194  
DB 273 LRHCSEILKEMAKKHFSYAMPFPYNDVNALGLANYVKNPKMLDGTIKEMDQEK 332  
QY 195 -AYMGTLFKGMADLIMSN-----PNYIPNISRIKDDVNPYPLASSKMFKAFRVS 247  
DB 333 DAYS-----FADVRLMFMNCYKYNPDHEVYTMARMLD-----VF 369  
QY 248 KSHSSKVKNFIFYKDLAVLYDKLSVSDSIDSESVFVYETSGTESLAKIKAH----- 302  
DB 370 ETHFSKIP--IEPVESMPLCIYIKTDIT-----ETTGRENTNEASSEGNSSD 413  
QY 303 -----EFFKRVLKLREKIS-----IAGSFONFYEKIESEKPESSPK 340  
DB 414 DSEDERVKRLAQEQLKAVHQOQLOVLSQVPPFKLNKKKSKKKKKK 463

RESULT 15  
ID 09KN45 PRELIMINARY: PRT: 1181 AA.  
AC 09KN45:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE ICMF-RELATED PROTEIN.  
GN VCA0120.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_Taxid=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Yamathavan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Ulterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae.";  
RL Nature 406:477-483(2000).  
DR EMBL; AE004353; AAF96034.1; -.  
DR TIGR; VCA0120; -.  
KW Complete proteome.  
SQ SEQUENCE 1181 AA; 134777 MW; 3EE262B2B0FE7949 CRC64;

Query Match 6.2%; Score 106; DB 2; Length 1181;  
Best Local Similarity 23.3%; Pred. No. 65;  
Matches 77; Conservative 54; Mismatches 118; Indels 82; Gaps 17;

QY 28 GSKRAPEGLVDLFAEL-----ARPPSTRDLDTNYVDVYS--GASG 68  
DB 866 GSKPAIYTEVLASVDELKSTLSIQDAPDYGMAALDITKARVNLN-ADPIYTLKRSSG 924  
QY 69 IYKPED-MVVDLGINMSVLLTPSARLQAYVKNVY--APAVYKSESKRYAG----- 117  
DB 925 LPRPLDSMAKLADESMYVY-----KQEAIRKLEVRWTFEDVYKTFQSKLAGRYFNPAS 978  
QY 118 --DTIIGRVLF--PS-----YSQSSAMTMPKPIFYGESGNOFLGGLDNIKTMK 167  
DB 979 NKDVALADPEAFAPNGTLDNFYNOQKMFIDE-NISVASDSDSAOSIIRREVLDQIKQAO 1037  
QY 168 EIVSYVSLGCEIDLEVLFE-----DMNGMEYAYMGTLKFGMADLIMSNP 214  
DB 1038 KIRAEAFNRKGIIDVGSFVSPEPLSLNNKRSYLVNDQGFAYSHGP--KENNELIMPNT 1094  
QY 215 -----NYIP--NISRIIKDDVNPYPLASSKMFKAFRVSKSHSKVKNFIYVK 262  
DB 1095 LRDSAVSKVTLIPTQTNMSPSRLQ---IQGPMAFFRLDDGDVVASAOTS--VDFKFIYD 1149  
QY 263 DLAVLYDKLSVSDSIDSESVKVYETSGT 293  
DB 1150 GGEMFY-RINAADANPTEFLKSEKLSKT 1179

Search completed: March 20, 2002, 00:00:37  
Job time: 2423 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 22:15:13 ; Search time 63.46 Seconds  
(without alignments)  
120.921 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719

Sequence: 1 MKRKAKSLFLSTVLFQAQ.....FQNFVEKIESKPESSPKN 341

Scoring table: BIOSDM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PTC05.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125.5	7.3	242	1	US-08-599-480-4
2	110	6.4	1657	3	US-09-057-570-2
3	110	6.4	1805	3	US-09-057-570-7
4	97.5	5.7	530	4	US-08-975-762-73
5	97.5	5.7	530	4	US-09-295-028-73
6	97.5	5.7	530	4	US-09-106-582-73
7	97.5	5.7	546	4	US-09-066-047-2
8	97.5	5.7	590	4	US-08-975-762-64
9	97.5	5.7	590	4	US-09-295-028-64
10	97.5	5.7	590	4	US-09-106-582-64
11	94	5.5	1382	3	US-09-057-570-4
12	93	5.4	607	3	US-08-486-099-116
13	93	5.4	607	3	US-08-484-223B-116
14	93	5.4	607	3	US-08-919-597-116
15	93	5.4	607	3	US-08-475-668A-116
16	93	5.4	607	3	US-08-485-551A-116
17	93	5.4	607	3	US-08-471-913A-116
18	93	5.4	607	4	US-08-485-264A-116
19	93	5.4	608	3	US-08-360-107A-126
20	92	5.4	679	1	US-08-441-139-5
21	91.5	5.3	654	4	US-09-090-793-2
22	91	5.3	1861	2	US-08-790-912-4
23	90	5.2	978	2	US-08-415-593-43
24	90	5.2	1186	1	US-08-485-568A-4
25	90	5.2	1186	1	US-08-357-698-6
26	90	5.2	1186	2	US-08-590-554A-4
27	90	5.2	1186	2	US-09-184-223-4

28	90	5.2	1186	5	PCT-US93-12682-6	Sequence 6, Appl1
29	88.5	5.1	1180	4	US-08-660-148-2	Sequence 2, Appl1
30	88.5	5.1	1212	4	US-08-660-148-5	Sequence 5, Appl1
31	87.5	5.1	627	1	US-08-703-947-2	Sequence 2, Appl1
32	87.5	5.1	877	1	US-08-072-574-12	Sequence 12, Appl1
33	87.5	5.1	877	3	US-08-486-270-12	Sequence 12, Appl1
34	87.5	5.1	877	3	US-08-367-264-12	Sequence 12, Appl1
35	87.5	5.1	1180	1	US-08-072-574-8	Sequence 8, Appl1
36	87.5	5.1	1180	1	US-08-486-270-8	Sequence 8, Appl1
37	87.5	5.1	1180	3	US-08-367-264-8	Sequence 8, Appl1
38	87.5	5.1	1212	1	US-08-072-574-10	Sequence 10, Appl1
39	87.5	5.1	1212	1	US-08-486-270-10	Sequence 10, Appl1
40	87.5	5.1	1212	3	US-08-367-264-10	Sequence 10, Appl1
41	87	5.1	678	1	US-08-844-085-2	Sequence 2, Appl1
42	87	5.1	715	2	US-08-484-993B-10	Sequence 10, Appl1
43	87	5.1	715	2	US-08-484-158B-10	Sequence 10, Appl1
44	87	5.1	715	2	US-08-484-596A-10	Sequence 10, Appl1
45	87	5.1	715	2	US-08-480-150A-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1  
; Sequence 4, Application US/08599480  
; Patent No. 5735459  
; GENERAL INFORMATION:  
; APPLICANT: Blanco, David R.  
; APPLICANT: Miller, James N.  
; APPLICANT: Lovett, Michael A.  
; APPLICANT: Champion, Cheryl I.  
; APPLICANT: Tempst, Paul J.  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF A  
; TITLE OF INVENTION: T. Pallidum RARE OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599,480  
; FILING DATE: 23-JAN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Learn, June M.  
; REGISTRATION NUMBER: 31,238  
; REFERENCE/DOCKET NUMBER: 07419/018001 (CIP of 016001)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 242 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-599-480-4

Query Match 7.3%; Score 125.5; DB 1; Length 242;  
Best Local Similarity 22.4%; Pred. No. 5.1e-05;  
Matches 55; Conservative 31; Mismatches 104; Indels 55; Gaps 8;  
QY 48 PSSRIIDLTVYVDV--YSGASGVKPEDMVVDGLNNW----SVLLTFSARLQAVKN 100

```
Db 21 PIAAORSKVNTQAFYIDDEFGAS-----EDOGL-AMRAAGSKFTIKGFPILKYFEG 70
OY 101 SVVAAYAVKSESKRYAGDTLLIGVRLFPYSYSSSAMIMP-----PKPIPFYSGESGNQFL 155
Db 71 MPOAVRMAGSWGOKKREARFETGVECKEENRQGNMIDLPTFKGSDYEIPLRGVSG---- 126
OY 156 GKGLIDNITKMEIKVSVYSLSGEYIDLEVLFEDMNGMEYASMGTLKFKGMADLIWSNPN 215
Db 127 -----PDVWVMGAGYQYSLEALVYDCRGVHTLLIGNLDPFQGMKNTLSVSVPT 173
OY 216 YIPNISRIIKDDVPNTPLASSKMFKAERKSKSHSKYKNIFFVKDLRLVYDKLSVI 275
Db 174 HLPQTS-----RYLGSAGHLSFVGFRJRTSPSERVDYF-----VFDDQFKALA 217
OY 276 DSDID 280
Db 218 NMHID 222
```

## RESULT 2

```
US-09-057-570-2
; Sequence 2, Application US/09057570
; Patent No. 6013266
; GENERAL INFORMATION:
; APPLICANT: Segers, Ruud P.A.M.
; APPLICANT: Frey, Joachim
; TITLE OF INVENTION: Live attenuated Actinobacillus
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 6013266el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/057,570
; FILING DATE: 09-APR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7400
; TELEFAX: (301) 948-9751
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1657 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-057-570-2
```

```
Query Match 6.4%; Score 110; DB 3; Length 1657;
Best Local Similarity 21.5%; Pred. No. 0.048;
```

```
Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;
```

```
OY 22 TDGLAEGSKRAEPGELVIDFAE---LARDPSSTR---IDLTVYVDYVYSGASGIYKPEDM 75
Db 860 TEALFNSTFKOSPENALYDLSEYLSFNDPTFMKRGILLISRYID--YAKAGGYE---- 913
OY 76 VVDIGINNMVLLTPSARLOAYVKNYSVA---PAVVSSESKRYAGDTLLIGVRLFPYSYS 131
Db 914 -----NW-----AATSNLTJARLREAGVIFAESTDLKGD----- 942
OY 132 QSSAMIMPFKIPIFYSGESGNQFL--GKGLIDNITKMEIKVSVYSLGEYIDLEVLFEDM 189
```

```
Db 943 EKNNTILGSOXKDNNLGSGAGDILLIGGES--NDTLGSGYGADYIFRSKHGQD--IYEDT 999
OY 190 NMEYAVSNGTLKFKGMADLIWSNPNYIPNISRIIKDDVPNTPLASSMRKARFVSVKS 249
Db 1000 NDNRRARDITLKF-----TDVNY-AEVFRVRDNDL-----MLGYYHDT 1038
OY 250 HSSKYNFIIFYVKDLRLVYDKLSVSDSDIDSESVFK---VYETSGTESLRKLAHETP 305
Db 1039 DSVTVKSFYSHD---YQDKELEFA-DRSITRDELKAGLHYGTGDGND---INDHADW 1091
OY 306 KRYVLKREKISTAESEF 322
Db 1092 DSILEGKGKDILIRGGY 1108
```

## RESULT 3

```
US-09-057-570-7
; Sequence 7, Application US/09057570
; Patent No. 6013266
; GENERAL INFORMATION:
; APPLICANT: Segers, Ruud P.A.M.
; APPLICANT: Frey, Joachim
; TITLE OF INVENTION: Live attenuated Actinobacillus
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 6013266el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/057,570
; FILING DATE: 09-APR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7400
; TELEFAX: (301) 948-9751
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-057-570-7
```

```
Query Match 6.4%; Score 110; DB 3; Length 1805;
```

```
Best Local Similarity 21.5%; Pred. No. 0.057;
```

```
Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;
```

```
OY 22 TDGLAEGSKRAEPGELVIDFAE---LARDPSSTR---IDLTVYVDYVYSGASGIYKPEDM 75
Db 1008 TEALFNSTFKOSPENALYDLSEYLSFNDPTFMKRGILLISRYID--YAKAGGYE---- 1061
OY 76 VVDIGINNMVLLTPSARLOAYVKNYSVA---PAVVSSESKRYAGDTLLIGVRLFPYSYS 131
Db 1062 -----NW-----AATSNLTJARLREAGVIFAESTDLKGD----- 1090
OY 132 QSSAMIMPFKIPIFYSGESGNQFL--GKGLIDNITKMEIKVSVYSLGEYIDLEVLFEDM 189
Db 1091 EKNNTILGSOXKDNNLGSGAGDILLIGGES--NDTLGSGYGADYIFRSKHGQD--IYEDT 1147
OY 190 NMEYAVSNGTLKFKGMADLIWSNPNYIPNISRIIKDDVPNTPLASSMRKARFVSVKS 249
```

Db 1148 NNDNRADDTLKE-----TDVNY-AEVKRRVNDL-----MLFGYHD 1186  
Qy 250 HSSKVNFIFFYKDLRVLYDKLSVSDSDSSEVFK-----VYETGTESTLRKKAHETP 305  
Db 1187 DSVTVKSPFSYSHVD---YGFQDKLEFA-DRSITRDELIRKAGLHVGTDGND---IKQNAOW 1239  
Qy 306 KRYLKAREKISTAECSF 322  
Db 1240 DSLEGGKGNDLIRGy 1256

RESULT 4  
US-08-975-762-73  
; Sequence 73, Application US/08975762  
; Patent No. 6207169  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/975,762  
; FILING DATE: 21-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAKI, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.439  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-622-4800  
; TELEFAX: 206-682-6031  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 530 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-975-762-73

Query Match 5.7%; Score 97.5; DB 4; Length 530;  
Best Local Similarity 22.8%; Pred. No. 0.15; Mismatches 101; Indels 75; Gaps 16;  
Matches 68; Conservative 54;

Qy 74 DMVY-----DLGINMSVLLTPSARLOAYVN-----SVAPAVYKSE 111  
Db 110 DMVYKFAHDLGR-----VGSNSLRSLIKNIRIFODANGVFQDEREYENVLADSGMTES 163  
Qy 112 S-----KRYAGDTILGVRLFPSPYSQSSAMIMPFKIPYSGESGNOFLGKGLIDNITKM 166  
Db 164 SYVNKIRNALPSTIL-MECLEFPNRAE-----LHIPYDALAKDVVLGL-LQHRVADI 213  
Qy 167 KEIKVSVSL-GVEI---DLEVLPEDMNGMEYAYSMGTLEKFG-----WADLIW 211  
Db 214 VEISSDAVDISGDSIDDELQKLFEE---QYKNSLNFPEYRSADYITMAEDDLADYIV 269  
Qy 212 SNPNIYIPNISRILKD--DVPNYPLA---SSKMRKAFRYSKSHSKVKNFIYVKDL-- 264

Db 270 SDQEVNVEIKNSLHDDRDVLNLVFTDKNEALAYKAOEKSPEELVSDAGYTIEDIAL 329  
Qy 265 -RVLYDKLSVSDSDSSEVFKYE---TSGETSLRKLKAHETFRVL-KLREKIS 316  
Db 330 NNISKDVLPGVNRVNVVALNEGVESEMFRRSYGVGHIMKVIKRIKHEITREDELEKLEKIS 387

RESULT 5  
US-09-295-028-73  
; Sequence 73, Application US/09295028  
; Patent No. 6277381  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.439C4  
; CURRENT APPLICATION NUMBER: US/09/295,028  
; CURRENT FILING DATE: 1999-04-20  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 73  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Ehrlichia sp.  
US-09-295-028-73

Query Match 5.7%; Score 97.5; DB 4; Length 530;  
Best Local Similarity 22.8%; Pred. No. 0.15; Mismatches 101; Indels 75; Gaps 16;  
Matches 68; Conservative 54;

Qy 74 DMVY-----DLGINMSVLLTPSARLOAYVN-----SVAPAVYKSE 111  
Db 110 DMVYKFAHDLGR-----VGSNSLRSLIKNIRIFODANGVFQDEREYENVLADSGMTES 163  
Qy 112 S-----KRYAGDTILGVRLFPSPYSQSSAMIMPFKIPYSGESGNOFLGKGLIDNITKM 166  
Db 164 SYVNKIRNALPSTIL-MECLEFPNRAE-----LHIPYDALAKDVVLGL-LQHRVADI 213  
Qy 167 KEIKVSVSL-GVEI---DLEVLPEDMNGMEYAYSMGTLEKFG-----WADLIW 211  
Db 214 VEISSDAVDISGDSIDDELQKLFEE---QYKNSLNFPEYRSADYITMAEDDLADYIV 269  
Qy 212 SNPNIYIPNISRILKD--DVPNYPLA---SSKMRKAFRYSKSHSKVKNFIYVKDL-- 264  
Db 270 SDQEVNVEIKNSLHDDRDVLNLVFTDKNEALAYKAOEKSPEELVSDAGYTIEDIAL 329  
Qy 265 -RVLYDKLSVSDSDSSEVFKYE---TSGETSLRKLKAHETFRVL-KLREKIS 316  
Db 330 NNISKDVLPGVNRVNVVALNEGVESEMFRRSYGVGHIMKVIKRIKHEITREDELEKLEKIS 387

RESULT 6  
US-09-106-582-73  
; Sequence 73, Application US/09106582  
; Patent No. 6306402  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:

THEAP

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,582  
FILING DATE: 29-JUN-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-622-6031  
INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 530 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-106-582-73

Query Match 5.7%; Score 97.5; DB 4; Length 530;  
Best Local Similarity 22.8%; Pred. No. 0.15;  
Matches 68; Conservative 54; Mismatches 101; Indels 75; Gaps 16;

QY 74 DMVV-----DLGNNMWSVLLPSPARLQAVKN-----SVVAPAVKSE 111  
DB 110 DMVVEKFAHDGIR-----VGSNSLRSLIKNIRIFDANGVFQDEREAVLADSGMES 163  
QY 112 S-----KRYAGDTLIGRVLPSPYSOSSAMIPPKIPYSESGNOPLGKGLINIKTM 166  
DB 164 SYVNKIRNALPSTIL-MECLFENRAE-----LHIPPYDALADVVLGL-LQHRVADI 213  
QY 167 KEIKSVSVSL-GYEI---DLEVLFEEDMNGMEYAVSMGTLKFKG-----WADLIW 211  
DB 214 VEISSDAVDISSDLSDELQKLFEE---QYKNSLNPEYRSADYIIMAEDDLADYIV 269  
QY 212 SNPNYIPNISRRIKD--DVNPYPLA---SSKMRKARFVSKSHSKYKNFTFYVKD-- 264  
DB 270 SPOEVDVEIKNSLHDQRYVNLVFTDKNEALAYKAVQEGKSFELVSDAGYTTIEDIAL 329  
QY 265 -RVLYDKLSVSDISDSISYFKYE---TSGTESLRLKLAHEFFKRYL-KLRKIS 316  
DB 330 NNISKDVLPGVGRNVVAFALNGEVSEMERSVVGMHIMKIRKHEITKEDLEKLEKIS 387

RESULT 7  
US-09-066-047-2  
Sequence 2, Application US/09066047A  
Patent No. 6306394  
GENERAL INFORMATION:  
APPLICANT: MURPHY, Cheryl  
STOREY, James  
BELTZ, Gerald A.  
COUGHLIN, Richard T.  
TITLE OF INVENTION: USE OF GRANULOCYTIC ERLHICHA  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: United States  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,047A  
FILING DATE: 24-Apr-1998  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/044,869  
FILING DATE: 25-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Superko, Colleen  
REGISTRATION NUMBER: 39,850  
REFERENCE/DOCKET NUMBER: 106,941.156  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 546 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-066-047-2

Query Match 5.7%; Score 97.5; DB 4; Length 546;  
Best Local Similarity 22.8%; Pred. No. 0.16;  
Matches 68; Conservative 54; Mismatches 101; Indels 75; Gaps 16;

QY 74 DMVV-----DLGNNMWSVLLPSPARLQAVKN-----SVVAPAVKSE 111  
DB 11 DMVVEKFAHDGIR-----VGSNSLRSLIKNIRIFDANGVFQDEREAVLADSGMES 64  
QY 112 S-----KRYAGDTLIGRVLPSPYSOSSAMIPPKIPYSESGNOPLGKGLINIKTM 166  
DB 65 SYVNKIRNALPSTIL-MECLFENRAE-----LHIPPYDALADVVLGL-LQHRVADI 114  
QY 167 KEIKSVSVSL-GYEI---DLEVLFEEDMNGMEYAVSMGTLKFKG-----WADLIW 211  
DB 115 VEISSDAVDISSDLSDELQKLFEE---QYKNSLNPEYRSADYIIMAEDDLADYIV 170  
QY 212 SNPNYIPNISRRIKD--DVNPYPLA---SSKMRKARFVSKSHSKYKNFTFYVKD-- 264  
DB 171 SPOEVDVEIKNSLHDQRYVNLVFTDKNEALAYKAVQEGKSFELVSDAGYTTIEDIAL 230  
QY 265 -RVLYDKLSVSDISDSISYFKYE---TSGTESLRLKLAHEFFKRYL-KLRKIS 316  
DB 231 NNISKDVLPGVGRNVVAFALNGEVSEMERSVVGMHIMKIRKHEITKEDLEKLEKIS 288

RESULT 8  
US-08-975-762-64  
Sequence 64, Application US/08975762  
Patent No. 6207169  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible

TREATM



Query Match	5.7%;	Score 97.5;	DB 4;	Length 590;
Best Local Similarity	22.8%;	Pred. No. 0.18;		

```

110 DMVVEKFAHDLGR-----VGSLSRLSLIKNFIQDANGVFOERYEAVLADSGMTES 163

```



Db 43 ASKAMEFSQSEDELTPFDMISIOQSVYISKIISDRDLAVVEYGTREKOKNSVN--FK 99  
Qy 162 NIKTMEIKVSVSLGYEIDL-----EVLPEMNGMEYVSMGTLFKGMADLIMSNPN 215  
Db 100 NIYVDELDPNGAKRILELDFQKGQOQKRFODMGHGSDSL-----SEVLWVCAN 151  
Qy 216 YIPNISRRI-----IKDDVP--NYPLASSKMRFA-----FRVSKSHSSKYNFI 258  
Db 152 LFSVDVFKMSHKRIMLFTNEDNPHGNDASAKASARTKAGDLRTGTGIFLDMHLKPGCF- 210  
Qy 259 FVYKDLRLVY-DKLSVSDSIDSESVFKYV--ETSGTES-LRKLKAHETFEKRV-----L 309  
Db 211 ----DISLFYRDIISIAEDEL-----RVHFEESSKLEDLRKVRKAKETRKRLSLRKL 260  
Qy 310 KLREKISIAEGSFQNFVEKIESEKP 334  
Db 261 KLNKDVIISVGIY-NLVQKALKPPP 284

## RESULT 13

US-08-484-223B-116  
; Sequence 116, Application US/08484223B  
; Patent No. 6020459  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Daniel P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
; NUMBER OF SEQUENCES: 245  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,223B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-029  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 116:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 607 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-484-223B-116

Query Match 5.4%; Score 93; DB 3; Length 607;  
Best Local Similarity 21.2%; Pred. No. 0.55;  
Matches 69; Conservative 53; Mismatches 117; Indels 86; Gaps 17;

Qy 43 ELARDPSTRDLTNVYDVYSGASGIVKPEDMYVDLGINNMSVLLTPSARLQAVKNSV 102  
Db 13 EEAEDEOEENLEASG--DYVYSGHDSLI-----FLVD----- 42  
Qy 103 VAPAVVSESKRYRAGDTILGVRLFPSPY-SQSSAMIMPPKIPYSGSGNOFLGKGLID 161  
Db 43 ASKAMEFSQSEDELTPFDMISIOQSVYISKIISDRDLAVVEYGTREKOKNSVN--FK 99  
Qy 162 NIKTMEIKVSVSLGYEIDL-----EVLPEMNGMEYVSMGTLFKGMADLIMSNPN 215  
Db 100 NIYVDELDPNGAKRILELDFQKGQOQKRFODMGHGSDSL-----SEVLWVCAN 151  
Qy 216 YIPNISRRI-----IKDDVP--NYPLASSKMRFA-----FRVSKSHSSKYNFI 258  
Db 152 LFSVDVFKMSHKRIMLFTNEDNPHGNDASAKASARTKAGDLRTGTGIFLDMHLKPGCF- 210  
Qy 259 FVYKDLRLVY-DKLSVSDSIDSESVFKYV--ETSGTES-LRKLKAHETFEKRV-----L 309  
Db 211 ----DISLFYRDIISIAEDEL-----RVHFEESSKLEDLRKVRKAKETRKRLSLRKL 260  
Qy 310 KLREKISIAEGSFQNFVEKIESEKP 334  
Db 261 KLNKDVIISVGIY-NLVQKALKPPP 284

## RESULT 14

US-08-919-597-116  
; Sequence 116, Application US/08919597  
; Patent No. 6054265  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Daniel P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
; NUMBER OF SEQUENCES: 273  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/919,597  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/470,896  
; FILING DATE: 06-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-020  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 116:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 607 amino acids  
; TYPE: amino acid







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 23:32:14 ; Search time 2311.81 Seconds  
(without alignments)  
7729.973 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663

Sequence: 1 atgataacttttttcaaaa.....atactctgaagaataatt 1663

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estda:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estcl:\*  
11: gb\_estc2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98.6	5.9	1101	13	CNS00EVL
2	94.6	5.7	1101	13	CNS00EVL
3	93.4	5.6	1101	13	CNS00EVL
4	88	5.3	928	13	CNS00EVL
5	86.8	5.2	1101	13	CNS00EVL
6	86.2	5.1	1029	13	CNS01ZGM
7	85.6	5.1	996	13	CNS00EVL
8	85.4	5.1	1101	13	CNS00EVL
9	85.2	5.1	1101	13	CNS00EVL
10	84.4	5.1	759	13	CNS00EVL
11	84	5.1	1101	13	CNS00EVL
12	83.6	5.0	945	13	CNS04DOK

C 13	83.6	5.0	1101	13	CNS0039G	AL063921 Drosophila
C 14	83.2	5.0	1043	13	CNS0145P	AL103735 Drosophila
C 15	82.8	5.0	1101	13	CNS0021J	AL061936 Drosophila
C 16	81.6	4.9	639	13	CNS038CX	AL233458 Tetradon
C 17	81.4	4.9	910	13	CNS01G8P	AL142826 Anopheles
C 18	81	4.9	1101	13	CNS00L72	AL078714 Drosophila
C 19	80.4	4.8	1092	13	CNS020K7	AL175696 Tetradon
C 20	80	4.8	1032	13	CNS020P	AL120674 Tetradon
C 21	79.6	4.8	1201	13	CNS0167M	AL106396 Drosophila
C 22	79	4.8	843	13	CNS0091L	AL053009 Drosophila
C 23	79	4.8	1101	13	CNS002FG	AL062437 Drosophila
C 24	78.8	4.7	905	13	CNS00KXK	AL077798 Drosophila
C 25	78.6	4.7	1101	13	CNS0008H	AL063632 Drosophila
C 26	78.6	4.7	1101	13	CNS058CV	AL325912 Tetradon
C 27	78.4	4.7	1101	13	CNS00E7W	AL106847 Drosophila
C 28	78.2	4.7	609	13	CNS025K2	AL182171 Tetradon
C 29	78.2	4.7	843	13	CNS00C51	AL059666 Drosophila
C 30	77.6	4.7	987	13	CNS014PQ	AL104456 Drosophila
C 31	77.6	4.7	1101	13	CNS00E07	AL069440 Drosophila
C 32	77.2	4.6	970	13	CNS0182A	AL108796 Drosophila
C 33	77.2	4.6	994	13	CNS04N0J	AL108972 Tetradon
C 34	77	4.6	836	13	CNS01100	AL099642 Drosophila
C 35	77	4.6	1027	13	CNS02T50	AL212733 Tetradon
C 36	77	4.6	1092	13	CNS020K7	AL175696 Tetradon
C 37	76.6	4.6	799	13	AZ184728	AZ184728 SP.1003.B
C 38	76.4	4.6	1061	13	CNS015LM	AL105604 Drosophila
C 39	76.4	4.6	1102	13	CNS014P4	AL104434 Drosophila
C 40	76.2	4.6	1101	13	CNS003BD	AL064082 Drosophila
C 41	76.2	4.6	1101	13	CNS034PQ	AL227373 Tetradon
C 42	76	4.6	661	13	CNS020V7	AL209800 Tetradon
C 43	75.8	4.6	668	10	AL514901	AL514901 Tetradon
C 44	75.8	4.6	668	10	AL514901	AL514901 Tetradon
C 45	75.8	4.6	1101	13	CNS017V2	AL108536 Drosophila

## ALIGNMENTS

RESULT 1  
CNS00EVL/c  
LOCUS  
DEFINITION  
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
AL069706  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr)

- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oseguwa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).





DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BRD89P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL069493
VERSION	AL069493.1 GI:4949636
KEYWORDS	GSS.
SOURCE	fruit fly
ORGANISM	Drosophila melanogaster
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)
TITLE	Genoscope.
JOURNAL	Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseguwa and Aaron Mamosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	Location/Qualifiers
Source	1..1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone.lib="RPCI-98"
	/clone="BACR29P01"
	/note="end : 77"
BASE COUNT	289 a 155 c 118 g 307 t 232 others
ORIGIN	
Query Match	5.6%; Score 93.4; DB 13; Length 1101;
Best Local Similarity	37.6%; Pred. No.0.00014;
Matches 198; Conservative 91; Mismatches 234; Indels 4; Gaps 1.	
OY	25 ttcttattcattctgcgaagagtgtgctgcatttaagaatatlaagaagatgaac 84
DB	1040 ttttttttttttdmrtgcttarrrrctgttatatkkatrrmmwtmdgatgacttggTTTTTT 981
OY	85 aaaattgaagataatttagacttccttiagtaaagaattcttcttiagaaacaatgag 144
DB	980 tTRTMDKAMAMWMTTTTKKTGWTMGWMTATKADRTDWMVTTADKTAATGAGCGAAM 921
OY	145 ggcttccttgaggcttatttaagaatgatlcaagaatttatacttcctttaagaalcy 204
DB	920 AKTTMAMTWMTATAKATTGTAGATKTCTGRRAATREDTTTDDAKAGCTTGATGCAGA 861
OY	205 gagcttatatgaactct---tcaagagagatttagatctcttataataagtggt 260
DB	860 KKGATTTTATITTAATAATDTAGMWATTAATTTAAAAAARAAWMTTTTWTATATMAA 801
OY	261 tataatgaactcttgatglttaagtlcaactctttctgcatgctcttiaataaacaaty 320
DB	800 TRRTTAAMAAMAMDRTAAAMAAATAATRTTTAAAAAAAWTATTTTNTAATAATATTW 741
OY	321 gccttaactaccctgyltaagalaaaaaaatatattatctctgtygaaggyglatcta 380
DB	740 AATWTMTWMTTTTTCGATAMWMAAAAAKAATATFRARAATAATATATATATATA 681
OY	381 aagggcgaaaagtgtaataatttactataaaaaataatlaaaaaaacgaatttt 440

Db	660	TWATTTTATTATTTTWTMTADNATTAATAATATATATATTTATTAATAAATAMNT	621
Oy	441	ataaaagattatatalaagaagtgttcaatgaagaagaactaaagatttat	500
Db	620	WAAAAATTTTARADATTAATAATTAAWAATTAATAAATAATTAATGAAATTTTMMTFA	561
Oy	501	tttttttatcacctgttccttttgccccaagaacgtagatla	547
Db	560	TATWTMTTWTTKTAAKTAAAATRTTAMATTAATTTATWTAMATTTTW	514

  

RESULT 4			
CNSDDKY	CNSODDKY	928 bp DNA	GSS 04-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC #		
DEFINITION	BACR27A24 of RPci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL071865		
VERSION	AL071865.1 GI:4948170		
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPc Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		

  

FEATURES	
source	location/Qualifiers
	1..928
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_1lb="RPci-98"
	/clone="BACR27A24"
	/note="end : 17"

  

BASE COUNT	
ORIGIN	
262 a	70 c 84 g 321 t 191 others

  

Query Match	5.3%;	Score 88;	DB 13;	Length 928;
Best Local Similarity	32.2%;	Pred. NO. 0.0009;		
Matches 118;	Conservative 100;	Mismatches 140;	Indels 0;	Gaps 0;

  

Oy	161	ttatttaagtgcatacaagatttatcatctttttaagaagtggattatcgtaa	220
Db	510	TTTTTTATKAWMAAAAAAATATTTTWTMTTWWMAAATAATTTTWT	569
Oy	221	tcttccaagagatltcatgatccttctaataaagtgatcataaatgactccttgatg	280
Db	570	TWATTTTATATTTATTTATTTAAAMWTAAATTTTATATWMAATTAATTTMTAT	629
Oy	281	taaggatcaacttttgtctatctcttaataaacaatgcgcttaactaccgtgtaa	340
Db	630	ATTAATATMTTWTMTATWTATATATATMAAAAAAATAAATAAATAAATTTTMA	689

Oy		341	gataaaaagattatcttctgtgaaggcggtactctaagaaggcaagtgttaaat	400
Db		650	AATTAATAAAAAAATTATTTTTTTTTTTTTTTWTATWAAATATAAATAATATMTATM	749
Oy		401	taaatttctaataaaaaataaaaaacgaataatttataaagaattatataag	460
Db		750	TWDCGKNMNMMNMAWWWWWMAWWMAAAMWWWWWMAAAAAAAAAAAAAAAAAAAAA	809
Oy		461	gagttggtttacctgaagaagcctaagaatttatcttttttatccactgt	520
Db		810	AWDDDDDDWDMDMWAKKKKKKKKKKKKKKBBKKKTXTTKTGARWTTTTTTTTTT	869
Oy		521	ctttt	526
Db		870	TTTTTT	875
RESULT		5		
CNSD0E07				
LOCUS		CNSD0E07	1101 bp	DNA GSS 04-JUN-1999
DEFINITION		Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR2P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION		AL069440		
VERSION		AL069440.1	GI:4949583	
KEYWORDS		GSS.		
SOURCE		fruit fly.		
ORGANISM		Drosophila melanogaster		
		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophila.		
REFERENCE		Genoscope.		
AUTHORS		Direct Submission		
TITLE		Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage		
JOURNAL		BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr)		
COMMENT		<p>- Web : www.genoscope.cns.fr )  determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and ESF libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.  Location/Qualifiers  1..1101  /organism="Drosophila melanogaster"  /db_xref="taxon:7227"  /clone_lib="RPCI-98"  /clone="BACR2P01"  /note="end : TET3"</p>		
FEATURES				
SOURCE				
BASE COUNT		366 a	66 c	104 g 351 t 214 others
ORIGIN				
Query Match		5.2%	Score 86.8;	DB 13; Length 1101;
Best Local Similarity		37.1%;	Pred. No. 0.0012;	
Matches 211; Conservative		84;	Mismatches 273;	Indels 1; Gaps 1;
Oy		15	tc metaaaggttttaccatctgcacgagggattgttgccaattaagatatra	74
Db		428	TKAAAAGAAARRGGCGKAATATTTGGAAGAATATATAAAAAAARATRAATTAATTA	487
Oy		75	agagatgacaanaattgaagaataatttagactctttagtaaagaatttctttaag	134

D	b		488	ATGATATATATAAAMTTTAAAAAAAMAAAATAAATAAACAAATWTAAATATTMTTAT	547
O	y		135	agcaatgaagggtcttccttcttggtcttttaataagtatcagaattttaatatcttc	194
D	b		548	AAATTAAMWMAKTCTTTTATWWATAATAAAAAAATTTATTTTTTATATATTWAA	607
O	y		195	ttaaagaatgaggttaataatgaacactccoaagaaatttaatgatctcttaataat	254
D	b		608	MGMWAAAAAAAMWTATRAAHHTTTAAATTTATATWTWMAATTTTAAATAMATTTATTAAM	667
O	y		255	ggtagataataagaactctttagatgyltaagtlcaactcltttgtatgcylctaala	314
D	b		668	TTAATWTTTTAAAWMTTAAWTATTAATTTAAAMWMAATTTWAAAAAMWMAATTTWMAAT	727
O	y		315	aacaatgaccttaactatccctglaagaataaaaaagatataatctctgttgaagg	374
D	b		728	AAAAATTAATTTTTTTTATTTAAAMWMTTAAAAATTTATATWAATAAAMWTTTATWMAAAAAA	787
O	y		375	atcttaagaagacaagltgataataatctaacttaataaaaataaataaaagaa	434
D	b		788	ATNTATTAAMATTTMKNKAATATATWTA-TAMWTTTAMATATWATRATAATTTATATTTT	846
O	y		435	aatttctaagaagatttatataaagaagctggttctactgtaaaaggaacttaaa	494
D	b		847	TWTTTATTTTAAATTTTTTTTAAWTAAAMWGTGAATWAAATTTAAWTTTAAAAAMAATGTT	906
O	y		495	tttaattcttttataatcacactctctcttgcctcaagagactgatatgaacag	554
D	b		907	TWTAAATTTWGTATWTTTATTTTAAATTTGATATATAAAAAAMWATTTAGATATWMAAA	966
O	y		555	gtcttaaaaggcagagcctgagaatta	583
D	b		967	TRTWTTGTATTKAAAATATWGRTATTKA	995

FEATURES	RESULTS
ACCESSION	1029 bp DNA
VERSION	GI:7812328
KEYWORDS	Genome survey sequence.
SOURCE	Tetraodon nigroviridis
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 1029)
AUTHORS	Roest-Crolius,H., Jalllon,O., Dasilva,C., Filzmes,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissensbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1029)
AUTHORS	Roest-Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Filzmes,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissensbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 1029)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/tetraodon">http://www.genoscope.cns.fr/tetraodon</a> .
FEATURES	Location/Qualifiers



[illegible]

```

Db      912 TTWTAATGATTGTCTTNAATGWAAAATAAATTTWTATAAAMATTTTAAATTAATTAATGAAGAAAAAMAATT   971
Oy      422 attaaagaacgaataatttataagaagattatatacttaaggagtgttcactgtaaaaag    481
Db      972 TWAATTATATATMAAAATTTTTTAATTAAATTTAAAWMAAAWMTAATTWATTNAAANAAAAAAAA          1031
Oy      482 aaagtaaagcatatttttttatcaccact    517
Db      1032 AWAANTAAANMKAAMWMAATTAAATTAATTAATTAATTTAAAT    1067

RESULT       9
CNS0021J     CNS0021J      1101 bp      DNA           GSS             03-JUN-1999
LOCUS        Drosophila melanogaster genome survey sequence TEI3 end of BAC #
DEFINITION   BACR05N11 of RPl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION    AL061936
VERSION      AL061936.1 GI:4940214
KEYWORDS     GSS.
SOURCE       fruit fly.
ORGANISM     Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    I (bases 1 to 1101)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqef@genoscope.cns.fr ; Web : www.genoscope.cns.fr)

COMMENT      Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html
              Aaron Mamoser in Pieter de Jong's laboratory at Kazutoyo Osoegawa and Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPl-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bdcpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source       1..1101
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone_1lb="RPl-98"
              /isogene="BACR05N11"
              /note="end : TEI3"
BASE COUNT   631 a                7 c            28 g            289 t            146 others
ORIGIN
Query Match 5.1%; Score 85.2; DB 13; Length 1101;
Best Local Similarity 41.2%; Pred. No. 0.0021;
Matches 219; Conservative 52; Mismatches 260; Indels 1; Gaps 1.;
Oy      60 aattaagatatttaagsagatgaacaatttgaaagataattgatcccttagtlaaaa    119
Db      450 AAAAAAAAAAAAAAAAANNANANAAAAAAAAAAAAAAAAAAAAAAAAANNAN         509
Oy      120 gatcttccttaagsgaacgaggggttccttccttgcgggttttctaagcgatccaaga    179
Db      510 AAAAATATAAATTTAATTTTWTATATTAAATTAATTTTWTATTTTWTATTAATTTTAA    569
Oy      180 tttaataatctcttttaagaatcgaggttatatgtcaaatccttcaagagaatttat    239
Db      570 TTTTAAANAANAATTTAATTAANAANWNTTTTAATTTTAATTAATTAANMAAAAAAAAAATT    629
```





```
DB 411 AAAMDNBNAAAAAAMNNNNNNYCATTT 387

RESULT 13
LOCUS CNS0039G/C
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BAC08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
GENOSCOPE.
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Oosawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source 1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BAC08K10"
/note="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN
Query Match 5.0%; Score 83.6; DB 13; Length 1101;
Best Local Similarity 14.3%; Pred. No. 0.0035;
Matches 70; Conservative 265; Mismatches 150; Indels 3; Gaps 1;
QY 25 tttttattcttctagcaagatgtgtgtctaatatgaatcttaagagatgaac 84
DB 1041 kmmmmwmtwtmwmkmmmmwmtakttdtamtwtamrwmagdragkrdrdaatdvg 982
QY 85 aaaaattgaaagataatagatcttttagtaaaagattctttagaagaagaagag 144
DB 981 aGRBDGGRKRRKDKRDKDDDDKGGKKKAKAAMKATKWWDDWMDKWDGAKDKK 922
QY 145 ggtttcttcttggttttcttaagtcaagattcttataatcttttttaagaagt 204
DB 921 ADDDDGAGKDDGKDKDADDTGDKDDDDKDKMDMDKAGTGWGDATWMAATDMMW 862
QY 205 gattatgatgtaattcttcaagagaattatgattctttaaataatggtatata 264
DB 861 GAAADADMT-...WDAADDMWADDRKDAWAMKWDMAAGARADRRDGRAGKRGAR 805
QY 265 atgaatcttctgattgaagtcattcttctgctatgctttaaataaacaatcgct 324
DB 804 KRDRKRADDKDAADDDDAATWTWTWTTRDTRDDKMKWKTDTWRMAADRWRDDDDD 745
```

```
QY 325 ttaactatccctgtaagataaaagattatcttcttgtaagggtatcttaag 384
DB 744 DRDRACTAGRRKRRKRRKRRDRFTWRDDADDTARDRRRRGDGAAGKRRKR 685
QY 385 agcaaatgtaataatattacttaataaaataataaagaaacgaataattataa 444
DB 684 RRDRATWDRTDKAWADAAWTTTDTDDMDKDRRRKRGARRRTTARAAMDMWTWA 625
QY 445 aagattatataaaggagttggtttacatgaagaagaaagcaagatattattt 504
DB 624 WDPAKWDWTRADRRWRMAADTWTDAKADBDWAKARARARADRARADRRWTKGR 565
QY 505 ttatatat 512
DB 564 TTTATWTT 557

RESULT 14
LOCUS CNS0145P/C
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL103735
VERSION AL103735.1 GI:5615346
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1043)
AUTHORS Direct Submission
TITLE Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billand at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
source 1. 1043
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN11G11"
/note="end : T7"
BASE COUNT 277 a 96 c 121 g 382 t 167 others
ORIGIN
Query Match 5.0%; Score 83.2; DB 13; Length 1043;
Best Local Similarity 38.9%; Pred. No. 0.0041;
Matches 165; Conservative 72; Mismatches 182; Indels 5; Gaps 1;
QY 140 tgaagggttcttcttggttttcttaagtgatcaagatttataatctttttaa 199
DB 959 KKKKKKKTTTTTTTTTTTTTTTTTTTATATATTTTDFRKKTTWDTATWGTTKGTA 900
QY 200 gaatgagttatagtaaatcttcaagagaatttatgattctttaaataatgta 259
DB 899 TRTDTTTTTTTTTTTTTTTTTTTTTT-...TTTTTTTTTTTTTTTTTTWTATWT 845
QY 260 ttaataatgattcttgatgaagtcattcttctgctatgctttaaataaacaat 319
DB 844 WATATATATWARRGGAADDAALAAATTTTWTWMMMMMMWTTTMMMMMMWTTTMM 785
```

OY	320	ggcgttcaacatccctgtaagataaaagaattatcattcttgtagagggtgatcct	379
Db	784	TTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAANATNTNMAATTATTAATAATWA	725
OY	380	aagaggacaagtgtgaataattctaataaaaataataaaaaacgaanaatt	439
Db	724	AAMTATMTATTAATAATATATMAATAAAAAATATTATWAAWWMAATTAATVTAWMAAAAAWA	655
OY	440	tataaagattatatataagaggttggtttacatgaagaagaaagctaaagatttta	499
Db	654	WAWMAAAAAMWTATATTTTWWMAWMTAAAMAWMAWMAAAAAAMWTATATMAATATTITTTWA	605
OY	500	tttttttatcatccactgltccttttcgtctoagaagactgtagattagaagggyltc	559
Db	604	TATATAAAAAATATATATAKTTTTTTTATATAMATAMWTATWTATATATWMAAAAAAGAGA	545
OY	560	aaaa 563	
Db	544	AAAA 541	

RESULT	15
CNS0021J/c	
LOCUS	
DEFINITION	CNS0021J 1101 bp DNA GSS 03-JUN-1999
DESCRIPTION	Drosophila melanogaster genome survey sequence TENG end of BAC # BACR05N11 of RPc1-9g library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION	AL061936	
VERSION	AL061936.1	GI:4940214
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila	melanogaster

REFERENCE  
1 (bases 1 to 1101)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

**JOURNAL**  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

**COMMENT**  
Determination of this BAC-end sequence was carried out as part of a

The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Oseagawa and Aaron Mammoso in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	Location/Qualifiers
source	1. .1101

BASE COUNT	ORIGIN
631 a	289 t
7 c	146 others
28 g	

Query Match	5.0%	Score 82.8	DB 13	Length 1101
Best Local Similarity	42.0%	Pred. No. 0.0046		
Matches 215; Conservative	46;	Mismatches 248;	Indels 3;	Gaps 1;

QY 18 aaaaaggttttatttcattctagcaaggattgttgctaatttaagatatttaaga 77

[illegible]

Search completed: March 20, 2002, 00:45:13  
Job time: 4379 sec



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 23:42:14 ; Search time 116.86 Seconds  
(without alignments) 3222.941 Million cell updates/sec

Title: US-09-004-395-1  
Perfect score: 1663  
Sequence: 1 atgataatcttttcaaaa.....atatacttgagaagatatatt 1663

Scoring table: IDENTITY-NUC  
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/lna/PCRTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/lna/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	4.7	19124	2	US-08-487-826B-13
2	63	3.8	665	2	US-08-883-795A-36
3	62.4	3.8	837	4	US-08-998-416-288
4	62.4	3.8	837	4	US-08-998-416-288
5	61.8	3.7	615	4	US-08-998-416-186
6	61.4	3.7	636	4	US-08-998-416-1137
7	59.6	3.6	615	4	US-08-998-416-186
8	59.2	3.6	636	4	US-08-998-416-1137
9	56	3.4	8920	2	US-08-446-855A-1
10	56	3.4	8920	2	US-09-150-741-1
11	55.4	3.3	4673	1	US-07-638-431-1
12	55.4	3.3	4673	1	US-07-638-431-1
13	55	3.3	19124	2	US-08-487-826B-13
14	54.4	3.3	665	2	US-08-883-795A-36
15	54	3.2	4818	3	US-08-817-925-27
16	53	3.2	6243	2	US-09-056-075-1
17	51.4	3.1	3600	3	US-08-894-731-1
18	51.2	3.1	1511	1	US-07-991-867B-8
19	51.2	3.1	1511	1	US-08-107-755A-8
20	51.2	3.1	1511	1	US-08-544-332-8
21	50.8	3.1	722	4	US-08-998-416-780
22	50.2	3.0	782	4	US-09-007-119-15
23	50.2	3.0	4810	3	US-08-852-629-11
24	50.2	3.0	4838	3	US-08-852-629-15
25	50	3.0	642	1	US-08-764-100-13
26	50	3.0	665	4	US-08-998-416-937
27	50	3.0	724	4	US-09-007-119-8

28	50	3.0	3000	1	US-08-764-100-9	Sequence 9, Appl1
29	50	3.0	6243	2	US-09-056-075-1	Sequence 1, Appl1
30	49.8	3.0	3095	6	5231168-1	Patent No. 5231168
31	49.8	3.0	51952	3	US-08-947-823-1	Sequence 1, Appl1
32	49.4	3.0	727	4	US-08-998-416-1011	Sequence 1011, Ap
33	49	2.9	660	4	US-07-991-867B-32	Sequence 32, Appl
34	49	2.9	660	4	US-08-107-755A-32	Sequence 32, Appl
35	49	2.9	660	2	US-08-544-332-32	Sequence 32, Appl
36	49	2.9	1511	1	US-07-991-867B-8	Sequence 8, Appl1
37	49	2.9	1511	1	US-08-107-755A-8	Sequence 8, Appl1
38	49	2.9	1511	2	US-08-544-332-8	Sequence 8, Appl1
39	49	2.9	1654	3	US-08-913-842-20	Sequence 20, Appl
40	49	2.9	1875	3	US-08-913-842-21	Sequence 1, Appl1
41	48.8	2.9	658	4	US-08-998-416-595	Sequence 595, App
42	48.8	2.9	4431	4	US-09-532-803-8	Sequence 8, Appl1
43	48	2.9	5852	1	US-07-867-106-2	Sequence 2, Appl1
44	47.8	2.9	319	1	US-07-593-657-14	Sequence 14, Appl
45	47.6	2.9	1172	1	US-07-945-288-9	Sequence 9, Appl1

## ALIGNMENTS

RESULT 1  
US-08-487-826B-13/c  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knodde Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned 29,655  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ. ID NO. 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13  
Query Match 4.7%; Score 78; DB 2; Length 19124;

Best Local Similarity 48.1%; Pred. No. 3.7e-07;  
Matches 222; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

```
OY 1 atgataaccttttccaagaagtttttatttcttctcagcaaggatttgccta 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15832 ATTATTTTATTTTAAATAAATTTTATTTTATGATATATTTTATTAACA 15773
OY 61 attaagatatttaagagatgaacaaatttgaaagaattagacttttagtaaa 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15772 TTTTATTTTAAATTTTATTTTATGATATATTTTATTTTATTTTCTT 15713
OY 121 attctcttgaagcaagaggttcttcttggttcttatttaagtgatcaag 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15712 TTTTATTTTATGATATATTTTATTTTATTTTATTTTATTTTCTTCT 15653
OY 181 ttctatactctttttaaagaatgaggttattatgtaaaccttccaagagaatt 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15652 TTTTATTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTT 15593
OY 241 attctttaaataatggttattataatcttctgactgaagtcactcttctg 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15592 AATAATCTTTTTCATTTTATTTATTCATCAAAATTTTATTTTATTTATTT 15533
OY 301 tgccttaataaaacaaatgcgcttcaactcctggaagataaaagattatct 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15532 TTTTAAATAATTTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 15473
OY 361 ttgttgaaggtgtaactttaaagagcaagttgataaatttacttaataaat 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15472 TATTTTCATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 15413
OY 421 aattaaaaacgaanaatttataaagaattatataaaga 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15412 AATACATATATATATAAATATATATATAAACAACATACAGTA 15371
```

## RESULT 2

```
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcive, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
; US-08-883-795A-36
```

Query Match 3.8%; Score 63; DB 2; Length 665;  
Best Local Similarity 47.8%; Pred. No. 0.00018;  
Matches 249; Conservative 0; Mismatches 265; Indels 7; Gaps 2;

```
OY 6 aatcttttcaaaaaggtttttatttcttctcagcaagggttgcctaatta 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33 AATATTTTATATTTAAATTTTATTTTAAATTTTATTTTATTTTATTTTATTT 92
OY 66 agatattaaagatgaacaaatttgaaagaatactttagactttagtaaaagatt 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93 AAATATTTTA--TAATTAATATTTTATTAATTAATTAATTTTATTAATTTTAT 150
OY 126 cttaagagaacatgaggggttcttcttggttcttattcaagtgatcaagatt 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 TAAATATTTTATTAATTAATTTTATTTTATTTTAAATTTTATTTTATTTAT 210
OY 186 tattctttttaaagaatgaggttattatgtaaaccttccaagaaattatgatt 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 TAAATATTTTATTAATTTATTTTATTAATTAATTTTATTAATTTTATTAAT 270
OY 246 tttaataatgltatlaaatgaalcttcttgatglttaaggtcaalcttctgcat 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 TAAATGTTTAAATTA--AAATTTTATTAATTTAAATGTTTATTAATTAATTA 325
OY 306 ttaataaaacaatgcgtttaaactatcctgtaagataaaaaagattattctgt 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 ATTAATTAATGTTTATTAATTTTATTTTATTTTAAATGTTTATTAATTAATTA 385
OY 366 gaaggtgtaactttaaagagcaagttgataaatttacttaataaaataaata 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 386 ATTAATTAATGTTTATTAATTTTATTAATTTTATTAATTTTAAATGTTTATTA 445
OY 426 aaaaacgaanaatttataaagattatataaagaaggttggtttaacagaaaga 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 ATTAATTAATGTTTATTAATTAATTTTATTAATTTTATTAATTTTAAATGTT 505
OY 486 cttaagattattattttttattatccactgttctttt 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 TTACATATTTTATTAATTAAGTATTTTATTAATTTTATTAATTTTATTTT 546
```

## RESULT 3

```
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschun, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264tn Carolina
; COUNTRY: USA
; ZIP: 27709
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
;
US-08-998-416-288

```

Query Match 3.8%; Score 62.4; DB 4; Length 837;

Best Local Similarity 45.1%; Pred. No. 0.00025; Matches 231; Conservative 0; Mismatches 281; Indels 0; Gaps 0;

```

QY 1 atgataatcttttcaaaaagggttttattcattcctcagaaggagttgtgcta 60
DB 83 ATTATATCATATTAATAAATTAATTTTCATTTATTAATCTATATATTAATTT 142
QY 61 attaagatacttaagaatgaacaaatttgaagaataattagacttttagtaaag 120
DB 143 ATATATATTTACTTAATTCATTTATTAATTTATATATTAATTAATAAATTAATTTAAT 202
QY 121 atttcttttaggaagcaatlgagggtttcttcttggttttatttaagtgtcaagat 180
DB 203 ATGAATACATTTAGTCTATGTTCAAAATTTTAAATTAGTTAATAATTAATTAAGATAT 262
QY 181 ttctaatctttttaaagaatlgagtttaattatgtaaaccttcaagagaattatcg 240
DB 263 TATTTATTTCTTTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTTA 322
QY 241 attccttataaatggtgataatgaatccttgatgtaaggtcaatcctttgcta 300
DB 323 TGTGTTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 382
QY 301 tctcttataaaaacacgcttaactaccctgtaagataaaaagaatataatc 360
DB 383 ATATATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTA 442
QY 361 tcttgaaagggtgatacttaaaggagaagtgataaactttaaactttaaactaaat 420
DB 443 TCTTTTATTAAGAATTAATTAATTAATTAATTTTAACTTAATTTCTTTATTAATTTT 502
QY 421 aattaaaaacgaacaaacttataaagaattatataaaggagtggttacaagaag 480
DB 503 TATATATTAATTAATAATTAATTAATTTCAATTTATTTATTTAATTAATTAATTAATTA 562
QY 481 gaaggctaaagatattatttttttattat 512
DB 563 TTTAATTAATTAATTTATTCATTTATTTAATTAAT 594

```

RESULT 4

```

US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippson, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
;
US-08-998-416-288

```

Query Match 3.8%; Score 62.4; DB 4; Length 837;

Best Local Similarity 52.2%; Pred. No. 0.00025; Matches 214; Conservative 0; Mismatches 186; Indels 10; Gaps 3;

```

QY 106 atcttttagtaagaattcttcttggaagcaatgagggtttcttcttggtttat 165
DB 656 ATATATATAGGTCGATCCCTTTATAGTATTTTAACTAATTCCTTAATAATATTTTAT 597
QY 166 taagtgatcaagaatttatacttctttaaagaatlgagtttaattagaacttt 225
DB 596 TAAATTAATTAATAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 537
QY 226 caagagaatttgaatcttcttaataatggtgataatgaatccttgatgtaag 285
DB 536 AATTAATAATGAATAA--TAAATTAATTAATAATTAATAATTAATAATTAATAATTAAG 479
QY 286 tcaactttttgtc----atgcttcaataaacaacatgcttcaactatcctgtgaag 341
DB 478 TTAATAATTAATTTAATAATTAATTTCTTAATAAAGATTAATAATTAATAATTAAT 419
QY 342 ataaaaagaattattcttcttgtaagggtgatactttaaaggagaagttgataaatt 401

```

Db 418 ATTATATAAATAGATATATTAATAAATAATATATTTACATATTTAA----ATAAAT 363  
Qy 402 aaatttactatataaataattataaagcaaaatttataaagatttataaagg 461  
Db 362 AAATCTTATATATATATAAATAATATTTATTTATTAACAAATTAATAATATTTAAT 303  
Qy 462 agttggttacaatgaagaagcctaaagattatttatttttttattat 511  
Db 302 TATGATATATCTATTTATATATTTATTAAGAAATAATATATATCTATATA 253

RESULT 5  
US-08-998-416-186  
; Sequence 186, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Redischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 186:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1074RP  
; US-08-998-416-186

Query Match 3.7%; Score 61.8; DB 4; Length 615;  
Best local similarity 47.8%; Pred. No. 0.00031;  
Matches 211; Conservative 0; Mismatches 227; Indels 3; Gaps 1;  
Qy 75 agagatgacaaatttgaagataattagatcttttaagtaaaagatttctttaga 134  
Db 10 ATATATATATTTAATTTAATAAGATTAAATTAACCTTTTATTAATAATTTAAGTATATA 69

Qy 135 agcaatgaggggtttcttcttgggttttatttaagtgatca---agatttattatct 191  
Db 70 AATATTTAAACCTATTTATTCATCTATTATTAATAATTAATTTGATTTATTAACCTTA 129  
Qy 192 ttttaagaatgaggttattatagtaaatcttccaagagattttatgattctttaat 251  
Db 130 TTATATATATTTATTTATTAATAATTTACTTAATTCATCATTTATTTATTAATAATAA 189  
Qy 252 aatggtgatttaaatgaatcttggatgtaaggtcaatccttttggatcttcaata 311  
Db 190 AATAATATTTATTAATGATTAATGATTAATTTACTGATGTCGAATTTTAATTAATTAATAA 249  
Qy 312 aaacaatgcgttactactccggtgaagataaaagattattatcttcttgaaagg 371  
Db 250 TATATATGATATATATATTTCTTTAATAATTAATTAATAAGATTATTAATAATA 309  
Qy 372 tgtatcttaagggcgaagtgtgataatlttaacttaataaaataaataaataaac 431  
Db 310 TATATTTTATTAATGTTGTTATTAATAATATATTTATTTATTAAGATTATTAATTTAT 369  
Qy 432 gaaatttataaagattatataaagaggttggtttacatgaaagaaagctaaaa 491  
Db 370 TTAATATGTAATTAATTAATTTATTTATTAATAATGATTTTATTAATAATTAATGAT 429  
Qy 492 gtaattattttttattat 512  
Db 430 TTATATATTTAATCTTTTAT 450

RESULT 6  
US-08-998-416-1137  
; Sequence 1137, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Redischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 1137:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 636 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1692RP  
US-08-998-416-1137

Query Match 3.7%; Score 61.4; DB 4; Length 636;  
Best Local Similarity 46.3%; Pred. No. 0.00038;  
Matches 243; Conservative 0; Mismatches 276; Indels 6; Gaps 1;

OY 1 atgataatcctttccaaaaaggcttttattcattcctcagcaaggagcttggtcta 60  
DB 83 ATTATATTCATTTTAAATTAATTAATTTGATTAATCACTATATTAATTAAT 142  
OY 61 atttaagatatttaagagatgacaaatttgaagataattagatcttttaagaag 120  
DB 143 ATATATTTACTTAAATTCATCTATTTATTAATTTATTAATTAATTAATTAAT 202  
OY 121 atttctcttgaagcaatgaggggttctcttcttggttttatttaagtgatcaagat 180  
DB 203 ATGAAATACATTTAGTCTATGTTCAAAATTTTAAATTAAGTTAAATTAATTAAT 262  
OY 181 ttttatactctttttaaagaatgagttataatgtaactc-----tcaagaagaat 234  
DB 263 TATTTATTTCTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 322  
OY 235 tttaagatcttttaataatggtgatttaataatccttcttgagtgtaagcgaactctt 294  
DB 323 TTGTTTATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAAT 382  
OY 295 ttgctatgctttaaacaacacgccttaactatccttgtaagataaaaaagatga 354  
DB 383 ATTATTAATTTTATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAAT 442  
OY 355 ttatctcgttggaagggtatcttaagaagcaagtgtaagaataatttactaat 414  
DB 443 TCTTTTATTAAGAAATTAATTAATTAATTTTAACTTAATTTCTTAATTTTAT 502  
OY 415 aaaaataatataaaaaacgaataattataagaattatataagaagcttggttact 474  
DB 503 TATATTTATTAATTAATTAATTAATTTTATTTATTTATTTATTTATTAATTAAT 562  
OY 475 gaaaaagaaagctaaagbatttatttcttataatccactgt 519  
DB 563 TTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 607

RESULT 7  
US-08-998-416-186/c  
; Sequence 186, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Bohmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jürgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reblschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ. ID NO.: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 3.6%; Score 59.6; DB 4; Length 615;  
Best Local Similarity 49.7%; Pred. No. 0.00085;  
Matches 181; Conservative 0; Mismatches 179; Indels 4; Gaps 1;

OY 148 ttctcttgggttttatttaagtgatcagaatttatactctttttaaagaatgag 207  
DB 612 TTCTTATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 553  
OY 208 ttattatgtaaaccttcaagaagaattatgattcctttaaataatggtatcaatg 267  
DB 552 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 493  
OY 268 aatccttgatgtaagcgaactcttctgctatgcttataataaacaatgcgctt 327  
DB 492 AATAAGAAATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 433  
OY 328 actatcctgtaagataaaaaagattatattcttgtaagggtgatcttaaaaggagc 387  
DB 432 TAAATCAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 373  
OY 388 aaagtgtataaatttaatttataaataaataaataaataaataaataaataaag 447  
DB 372 TAA---ATAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 317  
OY 448 atttataataaggaggtttacatgaaagaaagaaagaaagaaagaaagaaagaaag 507  
DB 316 AATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 257  
OY 508 acta 511  
DB 256 AATA 253

RESULT 8  
US-08-998-416-1137/c  
; Sequence 1137, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Bohmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jürgen





Db 434 AATATTATAATTCGTATTCATATTAAATTTGGAAATATTTTATAAGTTTATATTTCAGT 375  
 Oy 368 aggggtctactttaaaggagccaagcttgataaattaaatttbaactbaataaaataatbaa 427  
 Db 374 AATTTTATTAACACATGACACACAAACATTTATATATATATATATATATATATATAT 315  
 Oy 428 aaacgaaaattcttaaaagcttatatatataagagcttggtttcacatgaaaagaaagct 487  
 Db 314 ATATATATATTAATAATTAACCTTAAAGTATGTATTGTATATAATATATAAGAAAAA 255  
 Oy 488 aaagattcttacttttttttacta 511  
 Db 254 AAAAGTTTTTATCTATGTTATTA 231

```

US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbomoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-05-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

Query Match	3.48;	Score 56;	DB 4;	Length 8920;
Best Local Similarity	44.48;	Pred. No. 0.0074;		
Matches 224;	Conservative 0;	Mismatches 280;	Indels 0;	Gaps 0

OY	8	tcctttcccaaaaaaggtttcttcttccatcttcaggaaggttgctgctaattaaag	67
Db	734	TTTTTTTTTATACATTTATATTTATGTTTAAATATTTATTAATATTTACATATACAGTTC	675
OY	68	atattcaagagatgcacaaaattcgaaagaatcttgatcctttcagtaaaagattctt	127
Db	674	ATTTTCATATGTAAATTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTAA	615
OY	128	tttaggaagcaatgaagggtttccctttgggtcttcttaagatctcaagattcttata	187
Db	614	ATTACTACAATTACTATTTTATTAACATCAAGAAAAAATAATTAATGAATTAATTATA	555
OY	188	ttccttttcaagatgtagtcttcttgtaacctcttcaagaaattatgatctctc	247
Db	554	AATTAATATTAATTAATTAATATATGGAACCAATATTTGGGATATTAATATTTGTGA	409
OY	248	taataatgtagatataatgaatccttgcagttaagtcacatcttttgcabgtctc	307
Db	494	TAATATATAGGATGTTTATTATTAAGAAGGTAAACATATTAATGTGACACATCAA	435
OY	308	aataaaacaaatgcgctttaacttcctgtaagaataaaaaagatlaattcttggtga	367
Db	434	AATATTAATTAATTGATTCATATTAAATTTGGAATATATTAATAGTTTATATTTCAAGT	375
OY	368	agggtgctcttcaagggcgcaagtctgataataatcttaactcaataaaaaataatataa	427
Db	374	AATTTTATTAACCAATGACACCAACATTAATATATTAATATATTAATATATATAT	315

0y 428 aaagcaaaattcttaaaaggcttatcataagaagcttggttcacgaaagaagc 487  
Db 314 ATATTTATATAAATAACTTAAAGCTGATGTATTATTAATATAAGAAAAAAA 255

0y 488 aaaagcatcttactctttcttta 511  
Db 254 AAAAGTTTTTATCATGTATTAATA 231

```

RESULT 11
US-07-638-431-1/c
; Sequence 1, Application US/07638431
; Patent No. 5198535
;
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khushmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: Immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; CLONE: Py-lambdagt11-2-7 kb genomic expression
; LIBRARY: Py10.1111
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 718..3195
; OTHER INFORMATION:
;
US-07-638-431-1

```

Query Match	3.3%	Score 55.4;	DB 1;	Length 4673;
Best Local Similarity	45.7%	Pred. No. 0.0086;		
Matches 229; Conservative	0;	Mismatches 271;	Indels 1;	Gaps 1;

```
OY 1 atgataatcttttcaaaaaaggtttttatcttcaatctgaagaaggatttgcta 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3910 ATCATGAGCACTTGTGATATCTCTTTTAAACATTCATCTTTTTCACAC 3851
OY 61 attaagatatlaagaatgaacaaaattgaaagataattagatcctttagaaag 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3850 TTTTACTGATATAAACTTAAGACCATTAATTTATGTTGTAAATTTTAAATATCA 3791
OY 121 atttcttttagaagaagaagggttcttcttgggttttttaagtgatcaaat 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3790 CATTTTGTATTTCTTTTATCGATAAATGTGCGATTTGTCTATTTTAAAGAAAT 3731
OY 181 tttaatactttttaaagaatgagtttataatgtaaatcttcaagaagaattatg 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3730 TCATTTATATGATATATCATCTTTTGTATTTCATATTTTGTGAAATAGTAA 3671
OY 241 attccttaataatgagatataaagaatcctttagatgtaaggcaatcctttgcta 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3670 TACCGTGAATATACAAACCTAAATGTATTAACCTTTTATTTTATTTTAT 3611
OY 301 tgtcttaataaaaaaacaatgcgcttaactatcctgtaagataaaaaagtattatc 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3610 TTTTATTTATATTAAGATATTTGTAACCATTTAAATATTTGAGATATATATAT 3551
OY 361 ttgtgaagggtgatct-taaaggaagaagttgataaataatttactaaataaaa 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3550 ATTATACGACACATTTAGTTAAAAAAAATAGTAAAAAATCGTTAAAAAATA 3491
OY 420 taattaaaaaagaanaatttataaagattatataaagaagttggtttacatgaaa 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3490 TGAATAATTAATGATATATATTTGAATGAATACATTAATAAAGATACAAATTTATCAAACT 3431
OY 480 ggaagcctaaagaattttat 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3430 GTTAATTTAAATAACTTAAT 3410

RESULT 12
PCT-US92-00018-1/c
; Sequence 1, Application PC/7US9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedsstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: Immunogen and gene encoding
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMRC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00018
; FILING DATE: 19920103
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; LIBRARY: Py-lambda gtl1-2-7 kb genomic expression
; CLONE: Py10.1111
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 718..3195
; OTHER INFORMATION:
; PCT-US92-00018-1
```

```
Query Match 3.38; Score 55.4; DB 5; Length 4673;
Best Local Similarly 45.7%; Pred. No. 0.0086;
Matches 229; Conservative 0; Mismatches 271; Indels 1; Gaps 1;
```

```
OY 1 atgataatcttttcaaaaaaggtttttatcttcaatctgaagaaggatttgcta 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3910 ATCATGAGCACTTGTGATATCTCTTTTAAACATTCATCTTTTTCACAC 3851
OY 61 attaagatatlaagaatgaacaaaattgaaagataattagatcctttagaaag 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3850 TTTTACTGATATAAACTTAAGACCATTAATTTATGTTGTAAATTTTAAATATCA 3791
OY 121 atttcttttagaagaagaagggttcttcttgggttttttaagtgatcaaat 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3790 CATTTTGTATTTCTTTTATCGATAAATGTGCGATTTGTCTATTTTAAAGAAAT 3731
OY 181 tttaatactttttaaagaatgagtttataatgtaaatcttcaagaagaattatg 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3730 TCATTTATATGATATATCATCTTTTGTATTTCATATTTTGTGAAATAGTAA 3671
OY 241 attccttaataatgagatataaagaatcctttagatgtaaggcaatcctttgcta 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3670 TACCGTGAATATACAAACCTAAATGTATTAACCTTTTATTTTATTTTAT 3611
OY 301 tgtcttaataaaaaaacaatgcgcttaactatcctgtaagataaaaaagtattatc 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3610 TTTTATTTATATTAAGATATTTGTAACCATTTAAATTTGAGATATATATAT 3551
OY 361 ttgtgaagggtgatct-taaaggaagaagttgataaataatttactaaataaaa 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3550 ATTATACGACACATTTAGTTAAAAAAAATAGTAAAAAATCGTTAAAAAATA 3491
OY 420 taattaaaaaagaanaatttataaagattatataaagaagttggtttacatgaaa 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3490 TGAATAATTAATGATATATATTTGAATGAATACATTAATAAAGATACAAATTTATCAAACT 3431
OY 480 ggaagcctaaagaattttat 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3430 GTTAATTTAAATAACTTAAT 3410
```

```
RESULT 13
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
```



```

Db      154 TTTAATTTTAAATAATTTTAATTATAAATTTTAATTATAAATTTTAATTATAAATTTAT 15
Oy      333 tgataaattaaatttacttaataaaatlaaaaaacgaaaatttataaaagaltta 455
Db      94 TTTAATTTATAAAATATTTAATTATAAATATTTTAATTATAAATATTTTAATTATAAATA 35

Oy      453 tata 456
Db      34 TTTA 31

RESULT 15
US-08-817-926-27
; Sequence 27, Application US/08817926
; Patent No. 6001590
; GENERAL INFORMATION:
; APPLICANT: Komeda, Toshihiro
; APPLICANT: Suda, Hisako
; APPLICANT: Tamai, Yukio
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Kato, No. 6001590uo
; APPLICANT: Sakai, Yasuyoshi
; TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINI
; NUMBER OF INVENTION: FORMATE DEHYDROGENASE GENE
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,926
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/02597
; FILING DATE: 12-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 234133/1995
; FILING DATE: 12-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 42536/1996
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 081356/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4818 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Candida boidinii
; US-08-817-926-27

```

Query Match	3.28;	Score 54;	DB 3;	Length 4818;
Best Local Similarity	50.08;	Pred. No. 0.016;		

	Matches	162;	Conservative	0;	Mismatches	160;	Indels	2;	Gaps	1;
OY	173	ttcaagattttatactcttltttaagaagatggaattcataatgta--aaccttccaaga	230							
Db	4370	ttctgtgtttctttttctttttttattttctgttttttaatttttttttgatcaaaccttttaatt	4429							
OY	231	gaatttcgatctcttctaataatgtaagtaataatgaatccttttgatggttaagtcac	290							
Db	4430	ttagtttttaatt	4489							
OY	291	cttttgcatacgtccttaataaacaacatgcgcctttaaacttcctgtgaagaataaaaaag	350							
Db	4490	tttaatt	4549							
OY	351	attattatcttctgttgaagggtatctctaagaagsgcaagttgataaataaatttac	410							
Db	4550	tt	4609							
OY	411	taataaataataatbaaaaaacgaanaatttctaagaattcataataagaagtgtgttc	470							
Db	4610	ttttatt	4669							
OY	471	acatgaagaaggaagctaaagtta	494							
Db	4670	acgcataaaatgcataatataaacga	4693							

Search completed: March 20, 2002, 01:30:37  
Job time: 6503 sec



